SIMULATION-FREE SCHRÖDINGER BRIDGES ON RIEMANNIAN MANIFOLDS

Anonymous authors
Paper under double-blind review

ABSTRACT

Schrödinger bridges provide a principled framework for stochastic transport between distributions. Extending simulation-free Schrödinger bridge learning to Riemannian manifolds is challenging because neither Brownian transition kernels nor conditional bridge distributions are generally available in closed form. We introduce SSR-FM, a simulation-free framework for approximating Schrödinger bridges on Riemannian manifolds through flow and score matching. Using a small-noise heat-kernel expansion, SSR-FM constructs a geometry-aware approximation to the endpoint transport cost and obtains the corresponding entropic coupling by Sinkhorn. We then construct Gaussian conditional probability paths around minimizing geodesics, yielding tractable flow and score targets without simulating diffusion paths during training. Under regularity assumptions, we show that the reconstructed SDE reproduces the marginal distributions of the constructed probability path. Across benchmarks with analytically specified and data-estimated geometries, SSR-FM preserves manifold structure while achieving competitive hidden-time distribution matching among the existing methods.

1 INTRODUCTION

Learning stochastic transport between probability distributions is a central problem in generative modeling and in the inference of dynamics from temporal snapshots. In both settings, the marginal distributions alone do not determine how individual samples are transported between them (Tong et al., 2020; Lavenant et al., 2024). The Schrödinger bridge (SB) problem selects a stochastic process that connects the prescribed distributions while remaining close to a reference process (Léonard, 2014). Iterative SB methods learn these dynamics by estimating stochastic differential equation (SDE) drifts from simulated trajectories (De Bortoli et al., 2021; Shi et al., 2023), while flow matching and score matching provide tractable regression objectives for learning continuous-time dynamics (Song et al., 2021; Lipman et al., 2023; Tong et al., 2024a). In particular, SF²M combines an entropic transport coupling with directly sampled Brownian bridge marginals to learn SB dynamics without simulating diffusion paths during training (Tong et al., 2024b).

The geometry of the data matters when connecting distributions on curved or periodic spaces, such as geographic positions and molecular torsions. Riemannian Flow Matching (RFM) and Metric Flow Matching (MFM) incorporate geometry into conditional interpolation paths (Chen & Lipman, 2024; Kapuśniak et al., 2024). For stochastic transport, geometry also determines local diffusion and intermediate-state fluctuations. Riemannian diffusion models provide intrinsic stochastic dynamics (De Bortoli et al., 2022; Huang et al., 2022), while Riemannian Diffusion Schrödinger Bridge (RDSB) learns bridge dynamics through iterative SDE simulation (Thornton et al., 2022). Simulation-free learning on manifolds remains challenging because Brownian transition densities and conditional bridge distributions, which determine endpoint coupling and stochastic interpolation, generally lack closed forms. Although geodesic costs have been explored for endpoint coupling, existing approaches retain Euclidean Brownian bridge conditionals (Tong et al., 2024b). This motivates a simulation-free construction that derives both the coupling and conditional fluctuations from the same Riemannian Brownian reference.

We introduce SSR-FM, a simulation-free framework for learning Schrödinger bridges on Riemannian manifolds through flow matching and score matching. Starting from a given metric, SSR-FM uses the corresponding Riemannian Brownian motion as its reference process. To address the lack of

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

closed-form transition densities, we use a small-noise approximation to derive a transport cost that accounts for both geodesic separation and geometric volume distortion. The resulting entropic transport problem can be solved using Sinkhorn. We construct Gaussian conditional probability paths around minimizing geodesics, with covariance determined by the second variation of the action. These Gaussian distributions allow direct sampling of intermediate states for endpoint pairs drawn from the precomputed entropic coupling. They also yield tractable conditional flow and score targets, enabling regression-based training without simulating diffusion paths. The learned flow and score then define an SDE whose diffusion and coordinate drift retain the prescribed geometry. Thus, the same Brownian reference underlies the coupling, stochastic interpolation, and reconstructed dynamics.

Our main contributions are:

- We develop a simulation-free learning framework for Riemannian Schrödinger bridges, constructing approximate transport couplings and Gaussian conditional probability paths from a common Riemannian Brownian reference under both known and data-constructed metrics.
- We derive a geometry-aware coupling cost from a small-noise heat-kernel approximation and conditional Gaussian fluctuations determined by the Hessian of the Brownian action. We further show that the reconstructed SDE recovers the constructed probability path under ideal regression and regularity assumptions.
- We evaluate SSR-FM on simulated and real-world data with analytic metrics and metrics constructed from data. SSR-FM achieves lower mean distribution-matching errors than competing baselines on multiple benchmarks while maintaining geometric adherence on the prescribed manifolds.

2 Related Work

Flow matching and Schrödinger bridges. Score matching and its denoising formulation learn gradients of log densities, which support reverse-time diffusion models (Hyvärinen, 2005; Vincent, 2011; Song et al., 2021). Flow matching, rectified flow, and stochastic interpolants instead construct tractable probability paths and regress the fields that generate them (Lipman et al., 2023; Liu et al., 2023; Albergo & Vanden-Eijnden, 2023; Albergo et al., 2025). Coupling-based extensions use minibatch transport to improve the interpolations and their regression targets (Tong et al., 2024a; Pooladian et al., 2023). SBs select a path distribution relative to a reference process, with a classical decomposition into a static coupling problem and conditional reference bridges (Föllmer, 1988; Léonard, 2014; Chen et al., 2021). Neural approaches include iterative proportional fitting, forward–backward SDE training, and iterative Markovian fitting (De Bortoli et al., 2021; Chen et al., 2022; Shi et al., 2023). SF²M uses entropic couplings and direct Brownian bridge sampling for simulation-free flow and score matching (Tong et al., 2024b); LightSB and LightSB-M obtain tractable solvers through structured Schrödinger potentials and optimal bridge matching (Korotin et al., 2024; Gushchin et al., 2024). SSR-FM follows the conditional matching principle, but constructs its coupling and regression targets for a Riemannian Brownian reference rather than relying on Euclidean Brownian bridge formulas.

Generative modeling on Riemannian manifolds. Manifold normalizing flows account for geometry through intrinsic continuous-time dynamics, structured transformations, or direct likelihood-based maps (Mathieu & Nickel, 2020; Lou et al., 2020; Rezende et al., 2020; Sorrenson et al., 2024). RFM constructs conditional vector fields from premetrics, using analytic geodesics on simple geometries and spectral distances on more general domains (Chen & Lipman, 2024). MFM learns approximate geodesic interpolants under a data-dependent metric and includes both independent and OT-coupled variants (Kapuśniak et al., 2024). Riemannian diffusion models incorporate geometry into the noising process and score estimation (De Bortoli et al., 2022; Huang et al., 2022), with symmetric-space structure enabling more accurate transition-density calculations (Lou et al., 2023). RDSB extends iterative diffusion fitting to bridges on manifolds (Thornton et al., 2022). Related constructions only focus on simulation-free stochastic matching on $SE(3)$. (Bose et al., 2024). Our focus is a given Riemannian metric, whether known or constructed from data, that determines both the reference diffusion and the Gaussian conditional probability paths. The geometric construction therefore affects the coupling and conditional fluctuations as well as the interpolation mean.

Geometric transport and bridge approximations. Dynamic OT characterizes transport through evolving distributions and velocity fields, while entropic regularization makes discrete transport amenable to efficient matrix scaling (Benamou & Brenier, 2000; Cuturi, 2013; Peyré & Cuturi, 2019). Heat-kernel-based transport has been studied on geometric domains and manifold graphs (Solomon et al., 2015; Huguet et al., 2023), and the entropic cost associated with diffusion references has been analyzed in Riemannian settings (Conforti & Tamanini, 2021). The SF²M-Geo variant uses Geodesic Sinkhorn with a graph-based heat-kernel cost for endpoint coupling, while retaining Euclidean Brownian bridge conditionals (Tong et al., 2024b). Our approximation draws on the small-noise relation between transition densities and geodesic distance (Varadhan, 1967), short-time expansions accounting for geometric volume distortion (Ludewig, 2018; 2019), and Gaussian bridge fluctuations determined by the second variation along a minimizing path (Bailleul et al., 2021). These results provide the geometric coupling cost and conditional covariance used by SSR-FM. The method connects these approximations to explicit flow and score targets and to an SDE that preserves the marginals of the constructed probability path.

3 PRELIMINARIES

In this section, we consider stochastic transport between probability distributions on a Riemannian manifold with a given metric. We first introduce the Schrödinger bridge problem and its Riemannian Brownian reference (Section 3.1), then describe the connection to entropic optimal transport (Section 3.2).

3.1 RIEMANNIAN SCHRÖDINGER BRIDGES

The Schrödinger Bridge problem. Given a source distribution p_0, a target distribution p_1, and a reference stochastic process Q on $[0, 1]$, the Schrödinger bridge (SB) problem seeks a stochastic process P that bridges p_0 and p_1 and minimizes the Kullback–Leibler (KL) divergence from Q:

$$P^\star = \arg \min_{P:\, P_0=p_0,\, P_1=p_1} \mathrm{KL}(P\|Q). \tag{3.1}$$

Here, P_t denotes the marginal distribution of P at time t, and $\mathrm{KL}(P\|Q)$ is taken between the corresponding path distributions.

Riemannian Brownian motion. Let $(\mathcal{M}, g)$ be a smooth d-dimensional Riemannian manifold without boundary, with metric matrix $G(x) \succ 0$ in local coordinates. The choice of reference process determines how the bridge incorporates geometry. Riemannian Brownian motion provides a natural reference because its infinitesimal stochastic increments are isotropic with respect to the metric g. For $B_G(x)$ satisfying $B_G(x)B_G(x)^\top = G(x)^{-1}$, this process has the following local Itô representation:

$$\begin{aligned} \mathrm{d}X_t &= c_g(X_t)\,\mathrm{d}t + \sqrt{\varepsilon}\, B_G(X_t)\,\mathrm{d}W_t, \qquad X_0 \sim p_0, \\ c_g^i(x) &= \frac{\varepsilon}{2\sqrt{|G(x)|}} \sum_{j=1}^{d} \frac{\partial}{\partial x^j} \left(\sqrt{|G(x)|}\, g^{ij}(x) \right). \end{aligned} \tag{3.2}$$

Here, W_t is standard d-dimensional Brownian motion, $\varepsilon > 0$ is the noise variance rate, $g^{ij} = [G^{-1}]^{ij}$, and $|G| = \det G$. The noise has covariance $\varepsilon G^{-1}(X_t)\,\mathrm{d}t$, while c_g is the coordinate drift determined by the geometry (Appendix A.1). When $G = I$, $c_g = 0$ and the SDE reduces to $\mathrm{d}X_t = \sqrt{\varepsilon}\,\mathrm{d}W_t$.

Definition 3.1 (Riemannian Schrödinger bridge)**.** Given probability distributions p_0 and p_1 on a Riemannian manifold $(\mathcal{M}, g)$ and a reference stochastic process Q^g on $\mathcal{M}$, a Schrödinger bridge is a stochastic process $P^\star$ on $\mathcal{M}$ that satisfies

$$P^\star = \arg \min_{P:\, P_0=p_0,\, P_1=p_1} \mathrm{KL}(P\|Q^g). \tag{3.3}$$

In this work, we use the Riemannian Brownian motion in (3.2), initialized from p_0, as the reference process Q^g.

3.2 RIEMANNIAN ENTROPIC OPTIMAL TRANSPORT

The Schrödinger bridge can be obtained by solving a static problem over couplings of p_0 and p_1.

Proposition 3.2 (Schrödinger bridges via optimal couplings)**.** Assume that (3.3) admits a finite-KL solution. Let P be any feasible process with finite KL, and let $\pi = P_{01}$ and Q^g_{01} denote the joint distributions of (X_0, X_1) under P and Q^g, respectively. For $z = (x_0, x_1)$, write P^z and $Q^{g,z}$ for the corresponding conditional path distributions given $(X_0, X_1) = z$. Then

$$\mathrm{KL}(P\|Q^g) = \mathrm{KL}(\pi\|Q^g_{01}) + \mathbb{E}_{z\sim\pi}[\mathrm{KL}(P^z\|Q^{g,z})]\,. \tag{3.4}$$

For a fixed coupling π, the first term in (3.4) is fixed, while the second is nonnegative and vanishes when $P^z = Q^{g,z}$. Hence, among processes with coupling π, the mixture of reference bridges minimizes the path-space KL divergence. Optimizing the coupling then yields the Schrödinger bridge.

$$\pi^\star = \arg\min_{\pi\in\Pi(p_0,p_1)} \mathrm{KL}(\pi\|Q^g_{01}), \qquad P^\star = \int_{\mathcal{M}\times\mathcal{M}} Q^{g,z}\,\pi^\star(\mathrm{d}z), \tag{3.5}$$

where $\Pi(p_0, p_1)$ is the set of couplings with marginals p_0 and p_1 (Föllmer, 1988). The proof is given in Appendix A.3.

Transition density. Let $k^g_\varepsilon(x_0, x_1)$ denote the transition density of the Riemannian Brownian reference, defined by $Q^g(X_1 \in \mathrm{d}x_1 \mid X_0 = x_0) = k^g_\varepsilon(x_0, x_1)\,\mathrm{d}\mu_g(x_1)$. Here $k^g_\varepsilon(x_0, \cdot)$ is a conditional density with respect to Riemannian volume μ_g, with $\mathrm{d}\mu_g(x) = \sqrt{|G(x)|}\,\mathrm{d}x$ in local coordinates.

Proposition 3.3 (Riemannian entropic optimal transport)**.** Suppose p_0, p_1 admit densities with respect to μ_g and the relevant terms are finite. For $\varepsilon > 0$, let $c_\varepsilon(x_0, x_1) := -\varepsilon \log k^g_\varepsilon(x_0, x_1)$. Then the static Schrödinger problem (3.5) has the same minimizers as the following entropic optimal transport problem (Conforti & Tamanini, 2021).

$$\pi^\star = \arg\min_{\pi\in\Pi(p_0,p_1)} \left\{ \int_{\mathcal{M}\times\mathcal{M}} c_\varepsilon(x_0, x_1)\,\pi(\mathrm{d}x_0, \mathrm{d}x_1) + \varepsilon\,\mathrm{KL}(\pi\|p_0 \otimes p_1) \right\}. \tag{3.6}$$

The first term is the transport cost, and the second is the entropic regularization relative to the independent coupling $p_0 \otimes p_1$. The same parameter ε determines both the Brownian variance rate and the strength of entropic regularization. The proof is given in Appendix A.3.

4 Simulation-Free Schrödinger Bridges on Riemannian Manifolds

The above Section 3 expresses the Riemannian Schrödinger bridge as a mixture of reference bridges weighted by the optimal coupling $\pi^\star$. Computing this coupling and the conditional bridge marginals requires Brownian transition densities, which are generally unavailable in closed form on Riemannian manifolds. We next introduce our simulation-free method for approximating Riemannian Schrödinger bridges under a fixed metric G. Section 4.1 develops a tractable approximation to the transport cost, enabling the coupling to be computed using Sinkhorn. Section 4.2 constructs directly samplable conditional distributions, and Section 4.3 derives the flow and score matching objective.

4.1 Cost Approximation and Coupling

Unlike Euclidean Brownian motion, whose time-one transition density is $k_\varepsilon(x_0, x_1) = (2\pi\varepsilon)^{-d/2}\exp[-\|x_1 - x_0\|^2/(2\varepsilon)]$, the Riemannian Brownian reference generally has no closed-form transition density. For a general metric, the drift and diffusion coefficients in (3.2) vary with position; we therefore use a small-noise approximation.

Theorem 4.1 (Small-noise approximation of the transition density)**.** For two points x_0, x_1 connected by a unique shortest geodesic, let $e(x_0, x_1) := \frac{1}{2}d_g(x_0, x_1)^2$ be the minimum Brownian action over unit time, where d_g is the geodesic distance. Under the regularity assumptions in Appendix A.4, as $\varepsilon \downarrow 0$,

$$k^g_\varepsilon(x_0, x_1) = (2\pi\varepsilon)^{-d/2}\exp\left[-\frac{e(x_0, x_1)}{\varepsilon}\right] J(x_0, x_1)^{-1/2}\{1 + O(\varepsilon)\}, \tag{4.1}$$

where $J(x_0, x_1)$ is the volume Jacobian of the exponential map (Ludewig, 2018; 2019). Its definition is given in Appendix A.4.

Theorem 4.1 provides an estimate of the Riemannian Brownian transition density in terms of geodesic distance and geometric volume distortion. Combined with the Riemannian entropic OT formulation in Proposition 3.3, this estimate yields a tractable expression for the transport cost.

Proposition 4.2 (Tractable coupling cost for entropic OT on Riemannian manifolds)**.** Under the assumptions of Theorem 4.1, let $c_\varepsilon(x_0, x_1) = -\varepsilon \log k^g_\varepsilon(x_0, x_1)$ be the cost in the entropic coupling problem (3.6). Then, as $\varepsilon \downarrow 0$,

$$c_\varepsilon(x_0, x_1) = e(x_0, x_1) + \frac{\varepsilon}{2} \log J(x_0, x_1) + \frac{d\varepsilon}{2} \log(2\pi\varepsilon) + O(\varepsilon^2). \tag{4.2}$$

The leading term $e = d_g^2/2$ measures geodesic separation, while $\frac{\varepsilon}{2} \log J$ accounts for geometric volume distortion. The term $\frac{d\varepsilon}{2} \log(2\pi\varepsilon)$ is independent of (x_0, x_1) and can be omitted without changing the optimal coupling. The derivations are given in Appendix A.4; Appendix A.9 explains the large-deviation interpretation of the leading action cost and the role of positive noise.

We compute an approximate coupling $\widehat{\pi}$ using the Riemannian entropic OT formulation in Proposition 3.3, with cost $\widehat{c}_\varepsilon = e + \frac{\varepsilon}{2} \log J$. Given source samples x_i and target samples y_j, we form the cost matrix $C_{ij} = \widehat{c}_\varepsilon(x_i, y_j)$ and solve the empirical problem using Sinkhorn scaling (Sinkhorn, 1964), following its application to entropic OT by Cuturi (2013). This alternately rescales the rows and columns of $K_{ij} = \exp(-C_{ij}/\varepsilon)$ to match the prescribed marginal weights. The resulting coupling $\widehat{\pi}$ is held fixed during training, and pairs (x_i, y_j) are sampled with probabilities $\widehat{\pi}_{ij}$ to construct the conditional distributions in Section 4.2. The discrete formulation and updates are given in Appendix A.5.

4.2 Conditional Probability Paths on Riemannian Manifolds

With the coupling $\widehat{\pi}$ fixed, we next construct a conditional probability path $p_t(x \mid z)$ for each sampled pair $z = (x_0, x_1)$. We seek a family that allows direct sampling at arbitrary training times and admits tractable flow and score targets (Chen & Lipman, 2024; Tong et al., 2024b).

Reference conditional path. By Proposition 3.2, the Schrödinger bridge conditioned on z coincides with the corresponding Riemannian Brownian reference bridge. For $0 < t < 1$, its time-t density with respect to μ_g is $p^g_t(x \mid z) = k^g_{\varepsilon t}(x_0, x)\, k^g_{\varepsilon(1-t)}(x, x_1)/k^g_\varepsilon(x_0, x_1)$. At leading order in the local small-noise approximation of Theorem 4.1,

$$p^g_t(x \mid z) \propto [J(x_0, x)J(x, x_1)]^{-1/2} \exp\big[-\varepsilon^{-1}\{t^{-1}e(x_0, x) + (1-t)^{-1}e(x, x_1)\}\big]. \tag{4.3}$$

The proportionality factor is independent of x. The sum in the exponent is minimized at $\gamma_t(z)$ on the constant-speed minimizing geodesic from x_0 to x_1. A second-order expansion around $\gamma_t(z)$ motivates a Gaussian conditional probability path. The following proposition gives the corresponding covariance, with the derivation provided in Appendix A.6.

Proposition 4.3 (Gaussian conditional probability path)**.** Under the regularity assumptions of Theorem 4.1, let $\gamma_t(z)$ be the constant-speed minimizing geodesic from x_0 to x_1. For each fixed $t \in (0, 1)$, define, in the time-independent coordinates specified in Appendix A.6,

$$H_t(z) = \nabla^2_x \left[\frac{e(x_0, x)}{t} + \frac{e(x, x_1)}{1-t} \right] \Bigg|_{x=\gamma_t(z)}. \tag{4.4}$$

Then $H_t(z) \succ 0$, and we set $\Sigma_t(z) = H_t(z)^{-1}$. For $X_t \sim p^g_t(\cdot \mid z)$, the coordinate fluctuations satisfy $\varepsilon^{-1/2}(X_t - \gamma_t(z)) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \Sigma_t(z))$ as $\varepsilon \downarrow 0$ (Bailleul et al., 2021). The covariance calculation and proof are given in Appendix A.6.

Gaussian conditional probability path. Motivated by Proposition 4.3, we define the Gaussian coordinate density $\widehat{p}^{\,\mathrm{d}x}_t(x \mid z) = \mathcal{N}(x; \gamma_t(z), \varepsilon\Sigma_t(z))$, with direct sampler

$$X_t = \gamma_t(z) + \sqrt{\varepsilon}\, \Sigma_t(z)^{1/2} \xi, \qquad \xi \sim \mathcal{N}(0, I). \tag{4.5}$$

Here $\widehat{p}_t = \widehat{p}^{\,\mathrm{d}x}_t / \sqrt{|G|}$ is the density relative to Riemannian volume. The mean follows the metric geodesic, while $\varepsilon\Sigma_t$ describes the spread of intermediate states around it. Averaging these conditional densities over $\widehat{\pi}$ defines the marginal probability path $\widehat{p}_t(x) = \int \widehat{p}_t(x \mid z)\, \widehat{\pi}(\mathrm{d}z)$. This construction supplies training samples without simulating diffusion trajectories; Section 4.3 derives the flow and score targets for these conditional probability paths.

4.3 Learning Riemannian Schrödinger Bridges via Flow and Score Matching

Using the conditional probability path in Section 4.2, we define a joint flow and score matching objective (Tong et al., 2024b). The conditional flow and score derived below serve as regression targets for $v_\theta(t,x)$ and $s_\theta(t,x)$, respectively. For continuously differentiable γ_t and $\Sigma_t \succ 0$ on interior time intervals, the Gaussian path admits the explicit targets below. Their derivations are given in Appendix A.7.

Flow matching. For each fixed z, a probability-flow field satisfying $\partial_t \widehat{p}_t + \mathrm{div}_g(\widehat{p}_t u_t^\circ) = 0$ is

$$u_t^\circ(x \mid z) = \dot{\gamma}_t + \frac{1}{2}\dot{\Sigma}_t \Sigma_t^{-1}(x - \gamma_t). \tag{4.6}$$

The first term transports the mean along the geodesic, while the second accounts for the changing covariance. Dots denote time derivatives with z fixed.

Score matching. The score is the Riemannian gradient of the log density relative to μ_g. For the conditional Gaussian path, it is

$$\nabla_g \log \widehat{p}_t(x \mid z) = G(x)^{-1}\left[-\frac{1}{\varepsilon}\Sigma_t^{-1}(x - \gamma_t) - \frac{1}{2}\nabla_x \log|G(x)|\right]. \tag{4.7}$$

Here $\nabla_g = G^{-1}\nabla_x$, and the term involving $\nabla_x \log|G|$ corrects for the conversion from coordinate density to Riemannian volume.

Training objective. The conditional flow and score provide tractable regression targets for learning the marginal fields. Although these targets depend on the sampled pair Z, both networks take only (t, X_t) as input. Their joint objective measures regression errors in the Riemannian norm $\|w\|^2_{g(x)} = w^\top G(x) w$ (Chen & Lipman, 2024):

$$\mathcal{L}_{\text{SSR-FM}}(\theta) = \mathbb{E}\big[\|v_\theta - u_t^\circ(X_t \mid Z)\|^2_{g(X_t)} + \lambda(t)^2 \|s_\theta - \nabla_g \log \widehat{p}_t(X_t \mid Z)\|^2_{g(X_t)}\big]. \tag{4.8}$$

Here $v_\theta = v_\theta(t, X_t)$ and $s_\theta = s_\theta(t, X_t)$. The expectation is over $t \sim U(0,1)$, $Z \sim \widehat{\pi}$, and X_t sampled using (4.5). The positive weight $\lambda(t)$ controls the contribution of score matching.

The following theorem establishes the equivalence between this conditional objective and marginal flow and score matching.

Theorem 4.4 (Equivalence of conditional and marginal objectives)**.** Fix G, $\widehat{\pi}$, and the conditional probability path independently of θ. Let $u_t^\circ(x) = \mathbb{E}[u_t^\circ(x \mid Z) \mid t, X_t = x]$ be the marginal probability-flow field of $\widehat{p}_t$, and define

$$\mathcal{L}_{\text{marg}}(\theta) = \mathbb{E}_{t,\, X_t \sim \widehat{p}_t}\big[\|v_\theta - u_t^\circ(X_t)\|^2_{g(X_t)} + \lambda(t)^2 \|s_\theta - \nabla_g \log \widehat{p}_t(X_t)\|^2_{g(X_t)}\big]. \tag{4.9}$$

Both objectives use the same training-time distribution. Under the regularity conditions in Appendix A.8,

$$\nabla_\theta \mathcal{L}_{\text{SSR-FM}}(\theta) = \nabla_\theta \mathcal{L}_{\text{marg}}(\theta). \tag{4.10}$$

The proof is given in Appendix A.8 and follows the Flow Matching derivation (Lipman et al., 2023) and its joint flow/score extension (Tong et al., 2024b). Consequently, the conditional targets suffice to optimize marginal flow and score matching without evaluating the mixture density or its score.

5 SSR-FM Algorithm

Algorithm 1 summarizes the training procedure of SSR-FM. The transport coupling is first computed using Sinkhorn and kept fixed throughout training. Each iteration samples a pair from this coupling and directly draws an intermediate state at a randomly selected time. The flow and score networks are then updated using the corresponding conditional targets, without simulating diffusion paths.

Algorithm 1 SSR-FM training

Require: Source and target samples, metric G, variance rate $\varepsilon > 0$, and initial networks v_θ, s_θ.

$C_{ij} \leftarrow \widehat{c}_\varepsilon(x_i, y_j), \qquad \widehat{\pi} \leftarrow \mathrm{Sinkhorn}(C, \varepsilon).$

while training **do**

  $z \sim \widehat{\pi}, \quad t \sim \mathrm{Unif}(0,1), \quad \xi \sim \mathcal{N}(0, I).$

  $x \leftarrow \gamma_t(z) + \sqrt{\varepsilon}\,\Sigma_t(z)^{1/2}\xi.$

  $\ell(\theta) \leftarrow \|v_\theta(t,x) - u_t^\circ(x \mid z)\|^2_{g(x)} + \lambda(t)^2 \|s_\theta(t,x) - \nabla_g \log \widehat{p}_t(x \mid z)\|^2_{g(x)}.$

  $\theta \leftarrow \mathrm{Update}(\theta, \nabla_\theta \ell(\theta)).$

end while

return v_θ, s_θ.

The conditional targets are given by (4.6)–(4.7), and the discrete Sinkhorn problem is specified in Appendix A.5.

SDE reconstruction. After training, we combine the learned flow and score with the Riemannian Brownian reference. Starting from $X_0 \sim p_0$, samples are generated by solving the local Itô SDE

$$\mathrm{d}X_t = \left[c_g(X_t) + v_\theta(t, X_t) + \frac{\varepsilon}{2} s_\theta(t, X_t)\right] \mathrm{d}t + \sqrt{\varepsilon}\, B_G(X_t)\, \mathrm{d}W_t. \tag{5.1}$$

Here c_g and B_G are defined in (3.2); under the assumptions in Appendix A.8, ideal regression makes this SDE reproduce the time marginals $\widehat{p}_t$.

6 Experiments

We evaluate whether SSR-FM can recover unobserved distributions while respecting data geometry, using synthetic surfaces, observations on known manifolds, and single-cell snapshots. We compare against OT-CFM (Tong et al., 2024a), SF²M (Tong et al., 2024b), Metric FM (Kapuśniak et al., 2024), RFM (Chen & Lipman, 2024), and RDSB (Thornton et al., 2022). We measure distributional accuracy with unprojected Euclidean W_1 and geometric adherence with mean Euclidean distance to the prescribed manifold, $D_{\mathcal{M}}$. Experimental details and additional results are in Appendix B.

6.1 SSR-FM learns accurate interpolation on curved surfaces

Our first experiment tests interpolation when observations constrain only the endpoints. On a curved surface, a short route through the embedding space can connect the endpoints while leaving the surface. We transport a spiral source to an S-shaped target on a rolled saddle and a helicoid, using the metrics induced by their embeddings. The predicted distribution at the hidden time point $t = 0.5$ is compared with an independent numerical Brownian-SB solution, so evaluation does not reuse SSR-FM's conditional training targets.

Table 1: W_1 and distance-to-manifold errors at the hidden time point $t = 0.5$ (↓).

Method	Rolled saddle		Helicoid	
	W_1	$D_{\mathcal{M}}$	W_1	$D_{\mathcal{M}}$
OT-CFM	0.704	0.680	2.002	0.008
SF²M (SDE)	0.698	0.679	1.993	0.129
Metric FM	<u>0.136</u>	**0.000**	0.111	**0.000**
RFM	0.173	**0.000**	<u>0.091</u>	**0.000**
RDSB	0.778	**0.000**	1.541	0.007
SSR-FM (SDE)	**0.109**	**0.000**	**0.060**	**0.000**

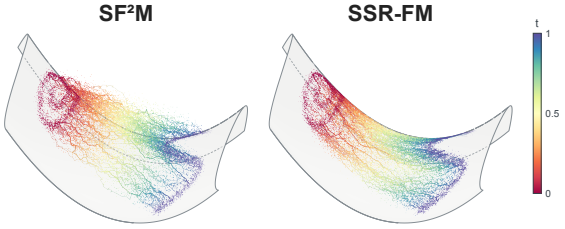


Figure 1: Rolled-saddle interpolation. Colors indicate time.

SSR-FM gives the lowest W_1 at the hidden time point on both surfaces (Table 1), reducing error relative to the strongest baseline by 19.7% and 34.7%, respectively. Compared with SF²M, it uses the surface metric to construct both the diffusion and conditional bridges, reducing off-surface motion

and distribution error (Figures 1 and B1). The improvement over Metric FM and RFM, which also stay on the surfaces, highlights the value of modeling stochastic spread as well as geometric paths. The matched Geodesic FM and covariance ablations in Appendix D further support the contribution of geometry-dependent conditional fluctuations to interpolation accuracy.

6.2 SSR-FM PRESERVES MANIFOLD GEOMETRY ON REAL-WORLD DATA

Next, we ask whether geometry-aware interpolation remains useful for real temporal observations. NOAA ocean drifters (Elipot et al., 2016) and IBTrACS cyclones (Knapp et al., 2010) lie on S^2, while ALA2 (Chodera et al., 2006) and Chignolin (Wang et al., 2019) use dihedral representations on T^2 and T^{18}. These tasks extend the comparison from geographic motion to periodic molecular configurations. We withhold two observation times and predict their distributions from the remaining unpaired snapshots, without supplying true trajectory correspondences. RFM, RDSB, and SSR-FM use the analytic geometry; Metric FM estimates a metric from data.

Table 2: Mean W_1 and distance-to-manifold errors over held-out time points on geographic and molecular datasets (↓). Geographic distances are in km; molecular distances use cosine–sine embedding units.

Method	NOAA (S^2)		IBTrACS (S^2)		ALA2 (T^2)		Chignolin (T^{18})	
	$W_1 \downarrow$	$D_{\mathcal{M}} \downarrow$	$W_1 \downarrow$	$D_{\mathcal{M}} \downarrow$	$W_1 \downarrow$	$D_{\mathcal{M}} \downarrow$	$W_1 \downarrow$	$D_{\mathcal{M}} \downarrow$
OT-CFM	108.418	5.192	126.555	9.917	0.319	0.138	**1.951**	0.618
SF²M (ODE)	102.102	6.744	<u>119.842</u>	8.273	0.398	0.077	2.099	0.765
SF²M (SDE)	102.323	7.801	125.202	24.965	0.351	0.074	2.105	0.763
Metric FM	113.207	12.279	129.900	14.193	0.629	0.292	2.097	0.765
RFM	<u>101.273</u>	**0.000**	122.775	0.022	0.316	**0.000**	2.354	**0.000**
RDSB	127.715	<u>0.003</u>	282.288	**0.010**	1.262	**0.000**	2.469	**0.000**
SSR-FM (ODE)	**101.228**	0.004	**119.722**	0.020	<u>0.314</u>	**0.000**	2.129	**0.000**
SSR-FM (SDE)	101.284	0.006	120.159	<u>0.011</u>	**0.308**	**0.000**	<u>1.985</u>	**0.000**

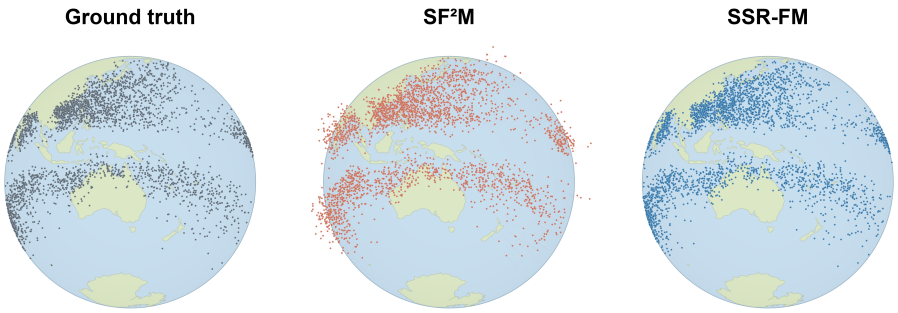


Figure 2: IBTrACS at the held-out 36 h time point. Radial deviations are magnified $10\times$ for display, not physical altitudes.

SSR-FM achieves the lowest mean W_1 on NOAA, IBTrACS, and ALA2, and the second-lowest on Chignolin (Table 2). On IBTrACS, SSR-FM achieves lower W_1 and much smaller manifold error than SF²M under both ODE and SDE sampling (Figure 2). RFM also respects these manifolds, yet SSR-FM reduces Chignolin W_1 from 2.354 to 1.985 with equally small manifold error. This comparison supports the value of geometry-aware endpoint coupling and stochastic bridge conditionals beyond constraining paths to the manifold. Together, the results show that SSR-FM improves the reconstructed distributions while retaining geometric fidelity.

6.3 SSR-FM RECONSTRUCTS SINGLE-CELL POPULATIONS WITH DATA-ESTIMATED GEOMETRY

Finally, we consider population data whose geometry is not available analytically. We use unpaired snapshots of epithelial–mesenchymal transition (EMT; Cook & Vanderhyden, 2020), embryoid-body differentiation (EB; Moon et al., 2019), and CITE-seq RNA (Burkhardt et al., 2022), withholding one intermediate population in each task. The goal is to reconstruct the missing population from the observed snapshots. The geometry-aware methods share a LAND metric estimated from training data (Arvanitidis et al., 2016); SSR-FM uses the scalable coupling and conditional computations described in Appendix C. This tests its applicability beyond manifolds with prescribed geometry. Results for all interior-time holdouts are reported in Appendix B.5.

Compared with SF²M under matched sampling modes, SSR-FM achieves lower W_1 on all six tasks (Table 3), supporting the use of data-estimated geometry in the reference diffusion. SSR-FM also outperforms Metric FM, which uses the shared LAND metric to shape interpolation paths. Our method additionally uses this metric to model stochastic fluctuations along its transport paths. These gains highlight the benefit of stochastic modeling under data-estimated metrics. Overall, SSR-FM obtains the best result on five tasks and is second to RFM on EB (50D), which demonstrates its advantage in single-cell population interpolation.

Table 3: W_1 errors for single-cell population interpolation on EMT, EB, and CITE-seq (↓).

Method	EMT (10D)	EB (10D)	EB (20D)	EB (50D)	EB (100D)	CITE (50D)
OT-CFM	2.1214	<u>1.5155</u>	3.1147	7.1398	11.5376	6.0854
SF²M (ODE)	2.1136	1.5181	3.1069	6.6484	10.6704	6.0717
SF²M (SDE)	2.1125	1.5179	<u>3.0958</u>	6.4114	<u>10.4378</u>	6.1131
Metric FM	2.1073	1.5580	3.6554	6.7708	11.2393	6.0527
RFM	2.1682	1.6987	3.1890	**6.1324**	10.4497	<u>6.0359</u>
RDSB	2.2699	2.9997	3.6453	6.5464	10.9956	6.3635
SSR-FM (ODE)	**2.0977**	1.5174	3.0991	6.5079	10.5303	**5.9968**
SSR-FM (SDE)	<u>2.1000</u>	**1.5091**	**3.0834**	<u>6.2795</u>	**10.3301**	6.0400

7 CONCLUSION

We introduced SSR-FM, a simulation-free framework for learning Schrödinger bridges on Riemannian manifolds through flow and score matching. Starting from a common Riemannian Brownian reference, SSR-FM incorporates geometry into both endpoint coupling and conditional stochastic interpolation. A small-noise approximation of the Riemannian Brownian reference yields a tractable coupling cost for Sinkhorn optimization and Gaussian conditional probability paths around minimizing geodesics. These paths provide directly samplable intermediate states and tractable regression targets, avoiding diffusion-path simulation during training. Under ideal regression and regularity assumptions, the reconstructed SDE recovers the marginals of the constructed probability path. Experiments on simulated and real-world datasets demonstrate strong interpolation performance with both known metrics and metrics constructed from data, together with geometric adherence on prescribed manifolds.

Limitations. SSR-FM relies on a local small-noise approximation around a single minimizing geodesic, which may lose accuracy at larger noise levels or when multiple geodesics contribute to the bridge. Computing geodesics and conditional covariances also introduces preprocessing costs, particularly for high-dimensional data-estimated metrics. Extending the approximation beyond a single geodesic and reducing these computational costs are important directions for future work.

AI USE STATEMENT

We used AI tools to assist with code implementation, literature searches, and identifying references relevant to the mathematical proofs. AI tools also assisted with manuscript drafting, translation, language editing, and proof checking. The authors independently reviewed all AI-assisted content and checked the proofs step by step. All final scientific and editorial decisions were made by the authors, who take full responsibility for the theory, mathematical results, and final content of this work.

REPRODUCIBILITY STATEMENT

Code, configurations, and reproduction instructions for the main experiments are available in our anonymous repository: https://anonymous.4open.science/r/Simulation-Free-Schrodinger-Bridges-on-Riemannian-Manifolds-581F. Appendix B documents the datasets, training settings, and evaluation protocols; Appendix A provides the mathematical derivations and proofs.

REFERENCES

Michael Albergo, Nicholas M. Boffi, and Eric Vanden-Eijnden. Stochastic Interpolants: A Unifying Framework for Flows and Diffusions. *Journal of Machine Learning Research*, 26(209):1–80, 2025.

Michael S. Albergo and Eric Vanden-Eijnden. Building Normalizing Flows with Stochastic Interpolants. In *International Conference on Learning Representations*, 2023.

Georgios Arvanitidis, Lars Kai Hansen, and Søren Hauberg. A Locally Adaptive Normal Distribution. In *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016.

Ismael Bailleul, Laurent Mesnager, and James Norris. Small-time fluctuations for the bridge of a sub-riemannian diffusion. *Annales Scientifiques de l'École Normale Supérieure*, 54(3):549–586, 2021. doi: 10.24033/asens.2465.

Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000. doi: 10.1007/s002110050002.

Avishek Joey Bose, Tara Akhound-Sadegh, Guillaume Huguet, Kilian Fatras, Jarrid Rector-Brooks, Cheng-Hao Liu, Andrei Cristian Nica, Maksym Korablyov, Michael Bronstein, and Alexander Tong. SE(3)-Stochastic Flow Matching for Protein Backbone Generation. In *International Conference on Learning Representations*, pp. 22590–22621, 2024.

Daniel Burkhardt, Malte Luecken, Andrew Benz, Peter Holderrieth, Jonathan Bloom, Christopher Lance, Ashley Chow, and Ryan Holbrook. Open problems - multimodal single-cell integration. `https://www.kaggle.com/competitions/open-problems-multimodal`, 2022. Kaggle.

Ricky T. Q. Chen and Yaron Lipman. Flow matching on general geometries. In *International Conference on Learning Representations*, 2024.

Tianrong Chen, Guan-Horng Liu, and Evangelos A. Theodorou. Likelihood Training of Schrödinger Bridge using Forward-Backward SDEs Theory. In *International Conference on Learning Representations*, 2022.

Yongxin Chen, Tryphon T. Georgiou, and Michele Pavon. Stochastic Control Liaisons: Richard Sinkhorn Meets Gaspard Monge on a Schrödinger Bridge. *SIAM Review*, 63(2):249–313, 2021. doi: 10.1137/20m1339982.

John D. Chodera, William C. Swope, Jed W. Pitera, and Ken A. Dill. Long-Time Protein Folding Dynamics from Short-Time Molecular Dynamics Simulations. *Multiscale Modeling & Simulation*, 5(4):1214–1226, 2006. doi: 10.1137/06065146x.

Giovanni Conforti and Luca Tamanini. A formula for the time derivative of the entropic cost and applications. *Journal of Functional Analysis*, 280(11):108964, 2021. doi: 10.1016/j.jfa.2021.108964.

David P. Cook and Barbara C. Vanderhyden. Context specificity of the EMT transcriptional response. *Nature Communications*, 11(1), 2020. doi: 10.1038/s41467-020-16066-2.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013.

Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion Schrödinger Bridge with Applications to Score-Based Generative Modeling. In *Advances in Neural Information Processing Systems*, volume 34, pp. 17695–17709. Curran Associates, Inc., 2021.

Valentin De Bortoli, Emile Mathieu, Michael Hutchinson, James Thornton, Yee Whye Teh, and Arnaud Doucet. Riemannian Score-Based Generative Modelling. In *Advances in Neural Information Processing Systems*, volume 35, pp. 2406–2422. Curran Associates, Inc., 2022. doi: 10.52202/068431-0175.

Shane Elipot, Rick Lumpkin, Renellys C. Perez, Jonathan M. Lilly, Jeffrey J. Early, and Adam M. Sykulski. A global surface drifter data set at hourly resolution. *Journal of Geophysical Research: Oceans*, 121(5):2937–2966, 2016. doi: 10.1002/2016jc011716.

Shane Elipot, Adam Sykulski, Rick Lumpkin, Luca Centurioni, and Mayra Pazos. Hourly location, current velocity, and temperature collected from global drifter program drifters world-wide, 2022. URL `https://www.ncei.noaa.gov/archive/accession/AOML-GDP-1hr`.

Hans Föllmer. Random fields and diffusion processes. In *École d'Été de Probabilités de Saint-Flour XV–XVII, 1985–87*, volume 1362 of *Lecture Notes in Mathematics*, pp. 101–203. Springer Berlin Heidelberg, 1988. doi: 10.1007/BFb0086180.

Jennifer Gahtan, Kenneth R. Knapp, Carl J. Schreck, III, Howard J. Diamond, James P. Kossin, and Michael C. Kruk. International best track archive for climate stewardship (IBTrACS) project, version 4.01, 2024. URL `https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.ncdc:C01552`.

Nikita Gushchin, Sergei Kholkin, Evgeny Burnaev, and Alexander Korotin. Light and Optimal Schrödinger Bridge Matching. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 17100–17122. PMLR, 2024.

Elton Hsu. *Stochastic Analysis on Manifolds*, volume 38 of *Graduate Studies in Mathematics*. American Mathematical Society, 2002. doi: 10.1090/gsm/038.

Chin-Wei Huang, Milad Aghajohari, Avishek Joey Bose, Prakash Panangaden, and Aaron Courville. Riemannian Diffusion Models. In *Advances in Neural Information Processing Systems*, volume 35, pp. 2750–2761. Curran Associates, Inc., 2022. doi: 10.52202/068431-0199.

Guillaume Huguet, Alexander Tong, María Ramos Zapatero, Christopher J. Tape, Guy Wolf, and Smita Krishnaswamy. Geodesic Sinkhorn For Fast and Accurate Optimal Transport on Manifolds. In *2023 IEEE 33rd International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–6. IEEE, 2023. doi: 10.1109/mlsp55844.2023.10285995.

Aapo Hyvärinen. Estimation of Non-Normalized Statistical Models by Score Matching. *Journal of Machine Learning Research*, 6(24):695–709, 2005.

Yuzuru Inahama. Large deviations for small noise hypoelliptic diffusion bridges on sub-riemannian manifolds. *Publications of the Research Institute for Mathematical Sciences*, 60(1):145–184, 2024. doi: 10.4171/PRIMS/60-1-4.

Erik Jørgensen. The central limit problem for geodesic random walks. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 32(1-2):1–64, 1975. doi: 10.1007/bf00533088.

Kacper Kapuśniak, Peter Potaptchik, Teodora Reu, Leo Zhang, Alexander Tong, Michael Bronstein, Avishek Joey Bose, and Francesco Di Giovanni. Metric Flow Matching for Smooth Interpolations on the Data Manifold. In *Advances in Neural Information Processing Systems*, volume 37, pp. 135011–135042. Curran Associates, Inc., 2024. doi: 10.52202/079017-4291.

Kenneth R. Knapp, Michael C. Kruk, David H. Levinson, Howard J. Diamond, and Charles J. Neumann. The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying Tropical Cyclone Data. *Bulletin of the American Meteorological Society*, 91(3):363–376, 2010. doi: 10.1175/2009bams2755.1.

Aleksandr Korotin, Nikita Gushchin, and Evgeny Burnaev. Light Schrödinger Bridge. In *International Conference on Learning Representations*, pp. 15354–15383, 2024.

Hugo Lavenant, Stephen Zhang, Young-Heon Kim, and Geoffrey Schiebinger. Toward a mathematical theory of trajectory inference. *The Annals of Applied Probability*, 34(1A):428–500, 2024. doi: 10.1214/23-aap1969.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow Matching for Generative Modeling. In *International Conference on Learning Representations*, 2023.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow. In *International Conference on Learning Representations*, 2023.

Aaron Lou, Derek Lim, Isay Katsman, Leo Huang, Qingxuan Jiang, Ser Nam Lim, and Christopher De Sa. Neural Manifold Ordinary Differential Equations. In *Advances in Neural Information Processing Systems*, volume 33, pp. 17548–17558. Curran Associates, Inc., 2020.

Aaron Lou, Minkai Xu, Adam Farris, and Stefano Ermon. Scaling Riemannian Diffusion Models. In *Advances in Neural Information Processing Systems*, volume 36, pp. 80291–80305. Curran Associates, Inc., 2023. doi: 10.52202/075280-3519.

Matthias Ludewig. Heat kernel asymptotics, path integrals and infinite-dimensional determinants. *Journal of Geometry and Physics*, 131:66–88, 2018. doi: 10.1016/j.geomphys.2018.04.012.

Matthias Ludewig. Strong short-time asymptotics and convolution approximation of the heat kernel. *Annals of Global Analysis and Geometry*, 55(2):371–394, 2019. doi: 10.1007/s10455-018-9630-4.

Christian Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete and Continuous Dynamical Systems*, 34(4):1533–1574, 2014. doi: 10.3934/dcds.2014.34.1533.

Emile Mathieu and Maximilian Nickel. Riemannian Continuous Normalizing Flows. In *Advances in Neural Information Processing Systems*, volume 33, pp. 2503–2515. Curran Associates, Inc., 2020.

Kevin R. Moon, David van Dijk, Zheng Wang, Scott Gigante, Daniel B. Burkhardt, William S. Chen, Kristina Yim, Antonia van den Elzen, Matthew J. Hirn, Ronald R. Coifman, Natalia B. Ivanova, Guy Wolf, and Smita Krishnaswamy. Visualizing structure and transitions in high-dimensional biological data. *Nature Biotechnology*, 37(12):1482–1492, 2019. doi: 10.1038/s41587-019-0336-3.

Gabriel Peyré and Marco Cuturi. Computational Optimal Transport with Applications to Data Sciences. *Foundations and Trends in Machine Learning*, 11(5-6):355–607, 2019. doi: 10.1561/2200000073.

Aram-Alexandre Pooladian, Heli Ben-Hamu, Carles Domingo-Enrich, Brandon Amos, Yaron Lipman, and Ricky T. Q. Chen. Multisample Flow Matching: Straightening Flows with Minibatch Couplings. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 28100–28127. PMLR, 2023.

Danilo Jimenez Rezende, George Papamakarios, Sebastien Racaniere, Michael Albergo, Gurtej Kanwar, Phiala Shanahan, and Kyle Cranmer. Normalizing Flows on Tori and Spheres. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 8083–8092. PMLR, 2020.

Simon Schwarz, Michael Herrmann, Anja Sturm, and Max Wardetzky. Efficient random walks on riemannian manifolds. *Foundations of Computational Mathematics*, 25(1):145–161, 2025. doi: 10.1007/s10208-023-09635-6.

Yuyang Shi, Valentin De Bortoli, Andrew Campbell, and Arnaud Doucet. Diffusion Schrödinger Bridge Matching. In *Advances in Neural Information Processing Systems*, volume 36, pp. 62183–62223. Curran Associates, Inc., 2023. doi: 10.52202/075280-2717.

Richard Sinkhorn. A Relationship Between Arbitrary Positive Matrices and Doubly Stochastic Matrices. *The Annals of Mathematical Statistics*, 35(2):876–879, 1964. doi: 10.1214/aoms/1177703591.

Justin Solomon, Fernando de Goes, Gabriel Peyré, Marco Cuturi, Adrian Butscher, Andy Nguyen, Tao Du, and Leonidas Guibas. Convolutional wasserstein distances: efficient optimal transportation on geometric domains. *ACM Transactions on Graphics*, 34(4):1–11, 2015. doi: 10.1145/2766963.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-Based Generative Modeling through Stochastic Differential Equations. In *International Conference on Learning Representations*, 2021.

Peter Sorrenson, Felix Draxler, Armand Rousselot, Sander Hummerich, and Ullrich Köthe. Learning Distributions on Manifolds with Free-Form Flows. In *Advances in Neural Information Processing Systems*, volume 37, pp. 66961–66994. Curran Associates, Inc., 2024. doi: 10.52202/079017-2138.

James Thornton, Michael Hutchinson, Emile Mathieu, Valentin De Bortoli, Yee Whye Teh, and Arnaud Doucet. Riemannian Diffusion Schrödinger Bridge. In *ICML Workshop on Continuous Time Methods for Machine Learning*, 2022.

Alexander Tong, Jessie Huang, Guy Wolf, David Van Dijk, and Smita Krishnaswamy. TrajectoryNet: A Dynamic Optimal Transport Network for Modeling Cellular Dynamics. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 9526–9536. PMLR, 2020.

Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, 2024a.

Alexander Y. Tong, Nikolay Malkin, Kilian Fatras, Lazar Atanackovic, Yanlei Zhang, Guillaume Huguet, Guy Wolf, and Yoshua Bengio. Simulation-free Schrödinger bridges via score and flow matching. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 1279–1287. PMLR, 2024b.

S. R. S. Varadhan. On the behavior of the fundamental solution of the heat equation with variable coefficients. *Communications on Pure and Applied Mathematics*, 20(2):431–455, 1967. doi: 10.1002/cpa.3160200210.

Juan Luis Vázquez. Asymptotic behaviour for the heat equation in hyperbolic space. *Communications in Analysis and Geometry*, 30(9):2123–2156, 2022. doi: 10.4310/CAG.2022.v30.n9.a7.

Pascal Vincent. A Connection Between Score Matching and Denoising Autoencoders. *Neural Computation*, 23(7):1661–1674, 2011. doi: 10.1162/neco_a_00142.

Jiang Wang, Simon Olsson, Christoph Wehmeyer, Adrià Pérez, Nicholas E. Charron, Gianni de Fabritiis, Frank Noé, and Cecilia Clementi. Machine Learning of Coarse-Grained Molecular Dynamics Force Fields. *ACS Central Science*, 5(5):755–767, 2019. doi: 10.1021/acscentsci.8b00913.

A Mathematical Derivations and Proofs

A.1 Coordinate Drift of Riemannian Brownian Motion

In this section, we derive the coordinate drift c_g in (3.2) by matching the generator of an Itô diffusion to $\varepsilon\Delta_g/2$.

Consider a local Itô diffusion with an unknown drift b and the prescribed covariance:

$$\mathrm{d}X_t^i = b^i(X_t)\,\mathrm{d}t + \sqrt{\varepsilon}\,B_G^{ia}(X_t)\,\mathrm{d}W_t^a, \qquad B_G B_G^\top = G^{-1}.$$

Repeated indices are summed, and ∂_i denotes differentiation with respect to x^i. For any smooth test function f, Itô's formula gives

$$\begin{aligned}\mathrm{d}f(X_t) &= \partial_i f\,\mathrm{d}X_t^i + \frac{1}{2}\partial_i\partial_j f\,\mathrm{d}[X^i, X^j]_t \\ &= \Big(b^i\partial_i f + \frac{\varepsilon}{2}g^{ij}\partial_i\partial_j f\Big)\,\mathrm{d}t + \sqrt{\varepsilon}\,\partial_i f\,B_G^{ia}\,\mathrm{d}W_t^a,\end{aligned} \tag{A.1}$$

where the coefficients and derivatives are evaluated at X_t. Riemannian Brownian motion with variance rate ε has generator $\varepsilon\Delta_g/2$ (Hsu, 2002). Expanding the Laplace–Beltrami operator yields

$$\begin{aligned}\frac{\varepsilon}{2}\Delta_g f &= \frac{\varepsilon}{2\sqrt{|G|}}\partial_j\Big(\sqrt{|G|}\,g^{ij}\partial_i f\Big) \\ &= \frac{\varepsilon}{2}g^{ij}\partial_i\partial_j f + \frac{\varepsilon}{2\sqrt{|G|}}\partial_j\Big(\sqrt{|G|}\,g^{ij}\Big)\,\partial_i f.\end{aligned} \tag{A.2}$$

Matching the $\mathrm{d}t$ coefficient in (A.1) with (A.2) for every f determines

$$b^i(x) = c_g^i(x) = \frac{\varepsilon}{2\sqrt{|G(x)|}}\partial_j\Big(\sqrt{|G(x)|}\,g^{ij}(x)\Big). \tag{A.3}$$

Therefore, the SDE in (3.2) describes Riemannian Brownian motion with generator $\varepsilon\Delta_g/2$. Since Δ_g is intrinsic, the resulting diffusion law is independent of the choice of coordinates. Appendix A.2 derives the same drift from isotropic tangent increments followed by geodesic steps, giving an equivalent construction in the continuous-time limit.

A.2 A Geodesic Construction of Riemannian Brownian Motion

In this section, we recover the Brownian drift in Appendix A.1 from isotropic tangent increments followed by geodesic steps. Throughout, $\varepsilon > 0$ is fixed and the time-step size h tends to zero. The calculation is local, with the construction stopped at the boundary of a relatively compact coordinate neighborhood.

Isotropic tangent increments. At the current point x, choose $B_G(x)$ whose columns form a g-orthonormal basis of $T_x\mathcal{M}$. Thus

$$B_G(x)^\top G(x) B_G(x) = I, \qquad B_G(x)B_G(x)^\top = G(x)^{-1}.$$

For a time step of length h, sample a fresh Brownian increment and form a tangent vector:

$$\Delta W \sim \mathcal{N}(0, hI), \qquad \eta = \sqrt{\varepsilon}\,B_G(x)\Delta W \in T_x\mathcal{M}. \tag{A.4}$$

The coefficients of η in this orthonormal basis are independent Gaussians with variance εh. Their distribution is unchanged by an orthogonal change of basis, so the tangent increment is isotropic with respect to g. In coordinates,

$$\mathbb{E}[\eta^i \mid x] = 0, \qquad \mathbb{E}[\eta^i \eta^j \mid x] = \varepsilon h\, g^{ij}(x). \tag{A.5}$$

From a geodesic step to coordinate drift. Let γ be the geodesic with $\gamma(0) = x$ and $\dot\gamma(0) = \eta$. For a step contained in the chart, take $x_{\mathrm{next}} = \gamma(1)$. The parameter $s \in [0,1]$ describes this single geodesic step; the corresponding increment in stochastic time is h. In local coordinates,

$$\ddot\gamma^i(s) = -\Gamma^i_{jk}(\gamma(s))\dot\gamma^j(s)\dot\gamma^k(s), \qquad \Gamma^i_{jk} = \frac{1}{2} g^{i\ell}\left(\partial_j G_{\ell k} + \partial_k G_{\ell j} - \partial_\ell G_{jk}\right).$$

Expanding the endpoint in the small initial tangent vector η gives

$$\begin{aligned} \Delta X^i := x^i_{\mathrm{next}} - x^i &= \dot\gamma^i(0) + \frac{1}{2}\ddot\gamma^i(0) + O(\|\eta\|^3_{g(x)}) \\ &= \eta^i - \frac{1}{2}\Gamma^i_{jk}(x)\eta^j\eta^k + O(\|\eta\|^3_{g(x)}). \end{aligned} \tag{A.6}$$

This is the standard second-order expansion of a geodesic step (Schwarz et al., 2025). The remainder is uniform on compact subsets of the chart. After localization, $\mathbb{E}[\|\eta\|^3_{g(x)} \mid x] = O(h^{3/2})$, so (A.5)–(A.6) imply

$$\begin{aligned} \mathbb{E}[\Delta X^i \mid x] &= -\frac{\varepsilon h}{2}\, g^{jk}(x)\Gamma^i_{jk}(x) + O(h^{3/2}), \\ \mathbb{E}[\Delta X^i \Delta X^j \mid x] &= \varepsilon h\, g^{ij}(x) + O(h^{3/2}). \end{aligned} \tag{A.7}$$

Dividing by h and taking $h \to 0$ yields covariance rate εG^{-1} and coordinate drift

$$c^i_g(x) = -\frac{\varepsilon}{2}\, g^{jk}(x)\Gamma^i_{jk}(x). \tag{A.8}$$

Thus the quadratic term of the geodesic expansion determines the coordinate drift of otherwise centered tangent increments.

Equivalence to the Itô form. To compare (A.8) with (A.3), use the inverse-matrix and determinant identities

$$\partial_j g^{ij} = -g^{i\ell} g^{jk} \partial_j G_{\ell k}, \qquad \partial_\ell \log|G| = g^{jk}\partial_\ell G_{jk}. \tag{A.9}$$

Expanding the Christoffel symbols and combining the first two terms by symmetry gives

$$\begin{aligned} -g^{jk}\Gamma^i_{jk} &= -g^{i\ell}g^{jk}\partial_j G_{\ell k} + \frac{1}{2} g^{i\ell} g^{jk}\partial_\ell G_{jk} \\ &= \partial_j g^{ij} + \frac{1}{2} g^{ij}\partial_j \log|G| \\ &= \frac{1}{\sqrt{|G|}}\partial_j\left(\sqrt{|G|}\, g^{ij}\right). \end{aligned} \tag{A.10}$$

Multiplying by $\varepsilon/2$ recovers precisely the drift in (A.3). Moreover, Taylor expansion of any smooth test function f and the moments in (A.7) give

$$\lim_{h\downarrow 0} \frac{\mathbb{E}[f(x_{\mathrm{next}}) \mid x] - f(x)}{h} = c^i_g(x)\partial_i f(x) + \frac{\varepsilon}{2} g^{ij}(x)\partial_i\partial_j f(x) = \frac{\varepsilon}{2}\Delta_g f(x). \tag{A.11}$$

The geodesic construction therefore has the same limiting generator as the Itô formulation in Appendix A.1. With the same initial distribution, both descriptions identify the same Brownian diffusion up to chart exit. The Brownian limit of geodesic random walks is classical (Jørgensen, 1975). More recently, Schwarz et al. (2025) establish convergence on compact manifolds for bounded isotropic steps combined with second-order retractions.

A.3 Static Schrödinger Problem and Entropic Transport

In this section, we prove Propositions 3.2 and 3.3 by reducing the path-space Schrödinger problem to an endpoint coupling and expressing its objective as entropic transport.

Path-space KL decomposition. For a path $X = (X_t)_{t\in[0,1]}$, write $z = (X_0, X_1)$. Conditional factorization gives

$$\log \frac{\mathrm{d}P}{\mathrm{d}Q^g}(X) = \log \frac{\mathrm{d}\pi}{\mathrm{d}Q^g_{01}}(z) + \log \frac{\mathrm{d}P^z}{\mathrm{d}Q^{g,z}}(X). \tag{A.12}$$

Taking expectation under P yields (3.4). The conditional KL term is nonnegative and vanishes when $P^z = Q^{g,z}$ for π-almost every z. Conversely, any coupling π with $\mathrm{KL}(\pi\|Q^g_{01}) < \infty$ defines a mixture $P = \int Q^{g,z}\,\pi(\mathrm{d}z)$ satisfying $P_{01} = \pi$ and $\mathrm{KL}(P\|Q^g) = \mathrm{KL}(\pi\|Q^g_{01})$. Minimizing over $\pi \in \Pi(p_0, p_1)$ proves (3.5).

Entropic transport objective. Write $p_1(x_1)$ for the target density with respect to μ_g. From the transition-density definition in Section 3.2 and $Q^g_0 = p_0$, the density ratio satisfies

$$\log \frac{\mathrm{d}\pi}{\mathrm{d}Q^g_{01}}(x_0, x_1) = \log \frac{\mathrm{d}\pi}{\mathrm{d}(p_0 \otimes p_1)}(x_0, x_1) + \log p_1(x_1) - \log k^g_\varepsilon(x_0, x_1). \tag{A.13}$$

Assuming the terms are integrable, integrating against π and using its fixed target marginal gives

$$\begin{aligned} \varepsilon\,\mathrm{KL}(\pi\|Q^g_{01}) = &\int_{\mathcal{M}\times\mathcal{M}} c_\varepsilon(x_0, x_1)\,\pi(\mathrm{d}x_0, \mathrm{d}x_1) \\ &+ \varepsilon\,\mathrm{KL}(\pi\|p_0 \otimes p_1) \\ &+ \varepsilon \int_{\mathcal{M}} p_1(x_1)\log p_1(x_1)\,\mathrm{d}\mu_g(x_1). \end{aligned} \tag{A.14}$$

The last term depends only on p_1, so removing it does not change the minimizing coupling. Since $\varepsilon > 0$, (3.5) and (3.6) have the same minimizers.

A.4 Small-Noise Transition Density and Transport Cost

In this section, we derive the small-noise transition density in Theorem 4.1 and the corresponding transport cost in Proposition 4.2.

Let $(\mathcal{M}, g)$ be a smooth, complete Riemannian manifold without boundary, with a non-explosive Brownian reference. We consider endpoint pairs (x_0, x_1) joined by a unique minimizing geodesic with non-conjugate endpoints, so the differential of the exponential map at its initial tangent vector is invertible. The estimates below are uniform on compact subsets of this regular-pair region, with t bounded away from zero.

We obtain the action and prefactor from the Fokker–Planck equation for the scalar transition density relative to μ_g. The short-time expansion for $L = -\Delta_g$ at time $\varepsilon t/2$ provides uniform estimates for the remainders and their derivatives (Ludewig, 2019).

Action from the density equation. Fix x_0 and write $\rho_t(y) = k^g_{\varepsilon t}(x_0, y)$. Relative to μ_g, Appendix A.1 gives

$$\partial_t \rho_t = \frac{\varepsilon}{2}\Delta_g \rho_t, \qquad \nabla_g f = G^{-1}\nabla_y f, \quad \mathrm{div}_g\, V = \frac{1}{\sqrt{|G|}}\partial_i(\sqrt{|G|}V^i), \quad \Delta_g f = \mathrm{div}_g(\nabla_g f). \tag{A.15}$$

For $I_\varepsilon = -\varepsilon \log \rho_t$, the chain and product rules give

$$\begin{aligned} \partial_t \rho_t &= -\frac{\rho_t}{\varepsilon}\partial_t I_\varepsilon, \\ \Delta_g \rho_t &= \frac{\rho_t}{\varepsilon^2}\|\nabla_g I_\varepsilon\|_g^2 - \frac{\rho_t}{\varepsilon}\Delta_g I_\varepsilon, \end{aligned} \qquad \partial_t I_\varepsilon + \frac{1}{2}\|\nabla_g I_\varepsilon\|_g^2 = \frac{\varepsilon}{2}\Delta_g I_\varepsilon. \tag{A.16}$$

The last identity is exact. For a curve on the unit interval, write its metric energy and Brownian action as

$$E_g(\gamma) = \int_0^1 \dot\gamma_s^\top G(\gamma_s)\dot\gamma_s \,\mathrm{d}s, \qquad S_g(\gamma) = \tfrac{1}{2}E_g(\gamma).$$

The small-noise leading term is the least action over duration t:

$$I(t, x_0, y) = \inf_{\gamma(0)=x_0,\,\gamma(t)=y} \frac{1}{2}\int_0^t \|\dot\gamma(s)\|_g^2 \,\mathrm{d}s = \frac{d_g(x_0,y)^2}{2t} = \frac{e(x_0,y)}{t}. \tag{A.17}$$

Indeed, Cauchy–Schwarz bounds the action below by the squared curve length divided by $2t$, and a constant-speed minimizing geodesic attains this bound. On the regular region, I solves $\partial_t I + \|\nabla_g I\|_g^2/2 = 0$. The cited short-time estimates give $\partial_t I_\varepsilon \to -e/t^2$, $\nabla_g I_\varepsilon \to \nabla_g e/t$, and a uniform bound on $\Delta_g I_\varepsilon$. Thus the right-hand side of (A.16) tends to zero; the two left-hand terms cancel in the limit. This identifies the exponential rate, but not a multiplicative prefactor. Appendix A.9 relates this minimum-action rate to large-deviation concentration; the prefactor requires the finer expansion below.

Geometric prefactor. Set

$$F_t(y) = (2\pi\varepsilon t)^{-d/2}\exp\left[-\frac{e(x_0,y)}{\varepsilon t}\right], \qquad A_t(y) = \rho_t(y)/F_t(y). \tag{A.18}$$

Write $e = e(x_0, y)$. Since $\|\nabla_g e\|_g^2 = 2e$, direct differentiation yields

$$\begin{aligned} \frac{\partial_t F_t}{F_t} &= \frac{e}{\varepsilon t^2} - \frac{d}{2t}, \\ \nabla_g F_t &= -\frac{F_t}{\varepsilon t}\nabla_g e, \\ \frac{\Delta_g F_t}{F_t} &= \frac{2e}{\varepsilon^2 t^2} - \frac{\Delta_g e}{\varepsilon t}. \end{aligned} \tag{A.19}$$

Using $\Delta_g(FA) = A\Delta_g F + 2\langle\nabla_g F, \nabla_g A\rangle_g + F\Delta_g A$, the two sides of (A.15), divided by F_t, are

$$\begin{aligned} \frac{\partial_t(F_t A_t)}{F_t} &= \left(\frac{e}{\varepsilon t^2} - \frac{d}{2t}\right)A_t + \partial_t A_t, \\ \frac{\varepsilon}{2F_t}\Delta_g(F_t A_t) &= \left(\frac{e}{\varepsilon t^2} - \frac{\Delta_g e}{2t}\right)A_t - \frac{1}{t}\langle\nabla_g e, \nabla_g A_t\rangle_g + \frac{\varepsilon}{2}\Delta_g A_t. \end{aligned} \tag{A.20}$$

Equating these expressions and cancelling the common $eA_t/(\varepsilon t^2)$ term gives

$$t\partial_t A_t + \langle\nabla_g e, \nabla_g A_t\rangle_g + \frac{1}{2}(\Delta_g e - d)A_t = \frac{\varepsilon t}{2}\Delta_g A_t. \tag{A.21}$$

The short-time expansion has $A_t = a_0 + O(\varepsilon t)$, with a_0 independent of t, and controls the derivatives in (A.21). Its leading equation is therefore

$$\langle\nabla_g e, \nabla_g a_0\rangle_g + \frac{1}{2}(\Delta_g e - d)a_0 = 0. \tag{A.22}$$

To define J, fix x_0 and express tangent vectors in a $g(x_0)$-orthonormal basis of $T_{x_0}\mathcal{M}$. Write $y = \mathrm{Exp}_{x_0}(\eta)$, where $\eta \in \mathbb{R}^d$ contains the coefficients of the initial tangent vector and y is the endpoint of its geodesic at parameter one. In the local coordinates used for $G(y)$, the volume Jacobian and change-of-variables formula are

$$J(x_0, y) := \sqrt{|G(y)|}\left|\det\left(\frac{\partial y}{\partial \eta}\right)\right|,$$
$$\mathrm{d}\mu_g(y) = J(x_0, y)\,\mathrm{d}\eta.$$

The derivative is taken with x_0 fixed, and $\mathrm{d}\eta$ is Euclidean volume in the orthonormal tangent coordinates. Setting $y = x_1$ gives $J(x_0, x_1)$ in (4.1).

In these normal coordinates, the volume density is J, $e = |\eta|^2/2$, and the components of $\nabla_g e$ are η^i by the first variation of geodesic energy. Consequently,

$$\Delta_g e = \frac{1}{J}\sum_i \frac{\partial(J\eta^i)}{\partial \eta^i} = d + \sum_i \eta^i \frac{\partial \log J}{\partial \eta^i}, \qquad \langle \nabla_g e, \nabla_g a_0\rangle_g = \sum_i \eta^i \frac{\partial a_0}{\partial \eta^i}. \tag{A.23}$$

Equation (A.22) becomes $\sum_i \eta^i \partial_{\eta^i}(a_0\sqrt{J}) = 0$, so $a_0\sqrt{J}$ is constant along each radial geodesic. Local unit-mass normalization gives $a_0(x_0, x_0) = 1$, and $J(x_0, x_0) = 1$; hence $a_0 = J^{-1/2}$ (Ludewig, 2018).

Substituting this coefficient and setting $t = 1$ gives (4.1), proving Theorem 4.1. Taking $-\varepsilon \log$ and using $\log(1 + O(\varepsilon)) = O(\varepsilon)$ gives (4.2), proving Proposition 4.2. The resulting $O(\varepsilon^2)$ statement concerns the cost after removal of its common constant; it is not a bound on the coupling error. The regular-branch assumptions exclude conjugate and cut-locus configurations.

A.5 Solving Schrödinger Bridges on Manifolds via Sinkhorn

In this section, we derive the discrete entropic transport formulation and Sinkhorn updates used to compute the approximate endpoint coupling $\widehat{\pi}$.

Let $C_{ij} = \widehat{c}_\varepsilon(x_i, y_j)$ be finite. With positive probability weights a, b, the discrete counterpart of (3.6) is

$$\begin{aligned} \widehat{\pi} = \arg\min_{\pi_{ij}\geq 0} \quad & \sum_{i,j} \pi_{ij} C_{ij} + \varepsilon \sum_{i,j} \pi_{ij} \log \frac{\pi_{ij}}{a_i b_j}, \\ \text{subject to} \quad & \sum_j \pi_{ij} = a_i, \qquad \sum_i \pi_{ij} = b_j. \end{aligned} \tag{A.24}$$

Strict convexity of the entropy gives a unique minimizing plan. Writing f_i, g_j for the marginal-constraint multipliers, stationarity at its positive entries gives

$$C_{ij} + \varepsilon\left(\log \frac{\pi_{ij}}{a_i b_j} + 1\right) + f_i + g_j = 0. \tag{A.25}$$

The factors depending only on i or j, including a_i, b_j, can be absorbed into positive scalings, giving $\widehat{\pi}_{ij} = \alpha_i K_{ij} \beta_j$ with $K_{ij} = e^{-C_{ij}/\varepsilon}$. Enforcing the two marginals alternately gives

$$\alpha_i \leftarrow \frac{a_i}{\sum_j K_{ij}\beta_j}, \qquad \beta_j \leftarrow \frac{b_j}{\sum_i K_{ij}\alpha_i}. \tag{A.26}$$

These are the standard Sinkhorn updates. For small ε, they are evaluated in the log domain with log-sum-exp. A common scalar factor in K is absorbed by the scalings and leaves $\widehat{\pi}$ unchanged, justifying removal of the constant in (4.2). The plan is fixed after convergence to the prescribed marginal tolerance.

A.6 Gaussian Conditional Probability Path

In this section, we prove Proposition 4.3 by expanding the Brownian bridge density around its minimizing geodesic. The quadratic term determines the local Gaussian covariance.

Density and coordinates. The Markov property gives the joint conditional density of $X_t = x$ and $X_1 = x_1$, given $X_0 = x_0$, as the product of the two transition densities. Dividing by the density of $X_1 = x_1$ gives the exact density ratio in Section 4.2. Multiplication by $\sqrt{|G(x)|}$ converts this density from Riemannian volume to coordinate Lebesgue measure.

For each pair z, fix a time-independent regular chart covering its minimizing geodesic. To define the full Gaussian sampler, take its coordinate image to be $\mathbb{R}^d$ and map coordinate samples back through the inverse chart. Density extensions and boundary behavior are required to satisfy the assumptions in Appendix A.8. The finite-noise Gaussian depends on this chart choice; replacing it by a truncated Gaussian or an exponential-map pushforward changes its density and targets.

Applying the two transition-density expansions locally around γ_t gives, up to factors independent of x,

$$p_t^{g,\mathrm{d}x}(x \mid z) \propto \frac{\sqrt{|G(x)|}}{\sqrt{J(x_0,x)J(x,x_1)}} \times \exp\left[-\frac{1}{\varepsilon}\left(\frac{e(x_0,x)}{t} + \frac{e(x,x_1)}{1-t} - e(x_0,x_1)\right)\right]\{1+O(\varepsilon)\}. \tag{A.27}$$

Expansion at the geodesic. Let $d_0 = d_g(x_0, x)$ and $d_1 = d_g(x, x_1)$. The weighted Cauchy–Schwarz inequality and the triangle inequality imply

$$\frac{e(x_0,x)}{t} + \frac{e(x,x_1)}{1-t} = \frac{1}{2}\left(\frac{d_0^2}{t} + \frac{d_1^2}{1-t}\right) \geq \frac{1}{2}(d_0+d_1)^2 \geq e(x_0,x_1). \tag{A.28}$$

Equality holds at $x = \gamma_t$, where $d_0 = t\, d_g(x_0, x_1)$ and $d_1 = (1-t)d_g(x_0, x_1)$. Uniqueness of the minimizing geodesic makes this the unique minimizer.

The first derivative also vanishes directly by the first variation of action. Reparameterize the two geodesic segments to unit time: $\gamma^L(s) = \gamma(ts)$ and $\gamma^R(s) = \gamma(t + (1-t)s)$. Their velocities at the joining point are $t\dot{\gamma}_t$ and $(1-t)\dot{\gamma}_t$. The coordinate derivatives of e with respect to the final and initial positions therefore give

$$\begin{aligned} \nabla_x e(x_0,x)|_{x=\gamma_t} &= t\, G(\gamma_t)\dot{\gamma}_t, \\ \nabla_x e(x,x_1)|_{x=\gamma_t} &= -(1-t)\, G(\gamma_t)\dot{\gamma}_t. \end{aligned} \tag{A.29}$$

Division by the respective durations cancels the two momenta, for any $t \in (0,1)$. Writing $\delta x = x - \gamma_t$, Taylor expansion gives

$$\frac{e(x_0,x)}{t} + \frac{e(x,x_1)}{1-t} - e(x_0,x_1) = \frac{1}{2}\delta x^\top H_t \delta x + O(\|\delta x\|^3). \tag{A.30}$$

The second variation of action is positive on nonzero variations with fixed initial and final positions along a non-conjugate minimizing geodesic. Its restriction to variations determined by the intermediate position is H_t, so $H_t \succ 0$.

Normalization. Retaining the quadratic term in (A.30) and replacing the smooth prefactor in (A.27) by its value at γ_t gives a constant times $\exp[-\delta x^\top H_t \delta x/(2\varepsilon)]$. The constant disappears on normalization, yielding

$$\widehat{p}_t^{\mathrm{d}x}(x \mid z) = \frac{\sqrt{\det H_t}}{(2\pi\varepsilon)^{d/2}} \exp\left[-\frac{(x-\gamma_t)^\top H_t (x-\gamma_t)}{2\varepsilon}\right]. \tag{A.31}$$

The change of variable $\xi = \varepsilon^{-1/2} H_t^{1/2}(x - \gamma_t)$ gives (4.5), with covariance $\varepsilon\Sigma_t = \varepsilon H_t^{-1}$. These are exact moments of the chosen Gaussian, not exact finite-noise moments of the Brownian bridge.

Small-noise convergence. Under the assumptions in Appendix A.4, the Brownian generator $\Delta_g/2$ is in symmetric divergence form, and completeness makes the unique minimizing geodesic strongly minimal. The centered, rescaled bridge is therefore tight and converges weakly to a Gaussian limit (Bailleul et al., 2021). A smooth extension of the coordinates outside a neighborhood of the geodesic does not change this limit, since the bridge concentrates on that neighborhood. Appendix A.9 explains why this concentration is compatible with retaining nonzero conditional fluctuations at positive noise.

For a fixed $t \in (0, 1)$, write $x = \gamma_t + \sqrt{\varepsilon}\, y$. The expansion (A.30) makes the exponent in (A.27) equal to $-\frac{1}{2} y^\top H_t y + O(\sqrt{\varepsilon}\,\|y\|^3)$, uniformly on bounded sets of y, while the smooth prefactor converges to its value at γ_t. Together with tightness, this identifies the normalized limiting density as $\sqrt{\det H_t}(2\pi)^{-d/2} \exp(-y^\top H_t y/2)$. Thus the limiting covariance is $H_t^{-1} = \Sigma_t$, proving Proposition 4.3. This fixed-time fluctuation limit does not assert that the finite-noise bridge is exactly Gaussian or provide a global KL error bound.

A.7 Conditional Flow and Score Targets

In this section, we derive the conditional flow and Riemannian score in (4.6)–(4.7) from the Gaussian conditional probability path.

Fix z and an interior interval $[a, b]$ on which γ_t and $\Sigma_t \succ 0$ are continuously differentiable. Write $\delta x = x - \gamma_t$ and suppress z in the density. Differentiating the Gaussian log density at fixed x gives

$$\begin{aligned} \partial_t \log \widehat{p}_t^{\mathrm{d}x} = & -\frac{1}{2}\operatorname{tr}(\Sigma_t^{-1}\dot{\Sigma}_t) + \frac{1}{\varepsilon}\dot{\gamma}_t^\top \Sigma_t^{-1}\delta x \\ & + \frac{1}{2\varepsilon}\delta x^\top \Sigma_t^{-1}\dot{\Sigma}_t\Sigma_t^{-1}\delta x. \end{aligned} \tag{A.32}$$

The spatial derivatives are $\nabla_x \widehat{p}_t^{\mathrm{d}x} = -\varepsilon^{-1}\widehat{p}_t^{\mathrm{d}x}\Sigma_t^{-1}\delta x$ and

$$\nabla_x^2 \widehat{p}_t^{\mathrm{d}x} = \widehat{p}_t^{\mathrm{d}x}\left[\varepsilon^{-2}\Sigma_t^{-1}\delta x\delta x^\top \Sigma_t^{-1} - \varepsilon^{-1}\Sigma_t^{-1}\right].$$

Multiplying (A.32) by $\widehat{p}_t^{\mathrm{d}x}$ and using these identities yields

$$\begin{aligned} \partial_t \widehat{p}_t^{\mathrm{d}x} &= -\dot{\gamma}_t^\top \nabla_x \widehat{p}_t^{\mathrm{d}x} + \frac{\varepsilon}{2}\operatorname{tr}(\dot{\Sigma}_t \nabla_x^2 \widehat{p}_t^{\mathrm{d}x}) \\ &= -\operatorname{div}_x\Big(\widehat{p}_t^{\mathrm{d}x}\dot{\gamma}_t - \frac{\varepsilon}{2}\dot{\Sigma}_t\nabla_x \widehat{p}_t^{\mathrm{d}x}\Big). \end{aligned} \tag{A.33}$$

Identifying the probability current in parentheses with $\widehat{p}_t^{\mathrm{d}x} u_t^\circ$ gives (4.6). Since the chart and metric do not vary with time,

$$\partial_t \widehat{p}_t^{\mathrm{d}x} + \operatorname{div}_x(\widehat{p}_t^{\mathrm{d}x} u_t^\circ) = \sqrt{|G|}\,[\partial_t \widehat{p}_t + \operatorname{div}_g(\widehat{p}_t u_t^\circ)] = 0. \tag{A.34}$$

Thus the same vector field generates the conditional density relative to μ_g. This is a valid choice of flow, not a uniqueness claim. No commutation of Σ_t and $\dot{\Sigma}_t$ is needed. In particular, this field need not equal the derivative of the symmetric-square-root sampler with a fixed noise realization.

For the score, $\widehat{p}_t = \widehat{p}_t^{\mathrm{d}x}/\sqrt{|G|}$ and (A.31) imply

$$\nabla_x \log \widehat{p}_t(x \mid z) = -\varepsilon^{-1}\Sigma_t^{-1}(x - \gamma_t) - \tfrac{1}{2}\nabla_x \log|G(x)|. \tag{A.35}$$

Multiplying by $G^{-1}(x)$ gives (4.7). The volume correction remains because the density used for sampling is a coordinate density. For computation, $\dot{\Sigma}_t = -\Sigma_t \dot{H}_t \Sigma_t$, obtained by differentiating $\Sigma_t H_t = I$. If pairs use different charts, conditional targets are converted to the same tangent representation before regression.

A.8 Conditional Regression and SDE Marginals

In this section, we prove the matching equivalence in Theorem 4.4 and the SDE marginal recovery result below. We follow the conditional-expectation argument of Flow Matching (Lipman et al., 2023), its Riemannian formulation (Chen & Lipman, 2024), and the joint flow/score extension (Tong et al., 2024b).

Theorem A.1 (Conditional regression and SDE marginals)**.** Fix a closed interval $[a, b] \subset (0, 1)$ and use the same training-time law for the conditional and marginal objectives. Let G, $\widehat{\pi}$, the conditional path, and the positive weight λ be independent of θ, and define the mixture density

$$\widehat{p}_t(x) = \int \widehat{p}_t(x \mid z)\, \widehat{\pi}(\mathrm{d}z). \tag{A.36}$$

Assume the following regularity conditions: spatial and temporal differentiation commute with integration against $\widehat{\pi}$; the targets and model outputs have finite metric second moments under the training law, with the stated score weight; and parameter differentiation commutes with expectation. Then the conditional and marginal objectives differ by a θ-independent constant and have identical gradients.

If, in addition, the training-time law has positive density on $[a, b]$, the model class can represent both conditional means, and the population minimum is attained with continuous representatives of the optimal fields, minimizing (4.8) gives

$$\begin{aligned} v_{\theta^\star}(t, x) &= \mathbb{E}[u_t^\circ(x \mid Z) \mid t, X_t = x], \\ s_{\theta^\star}(t, x) &= \nabla_g \log \widehat{p}_t(x), \\ \partial_t \widehat{p}_t + \mathrm{div}_g(\widehat{p}_t v_{\theta^\star}) &= 0. \end{aligned} \tag{A.37}$$

If the reconstructed SDE is non-explosive and its Fokker–Planck initial-value problem is unique in the considered probability class, with boundary flux or decay conditions permitting integration by parts, then (5.1), initialized with $\widehat{p}_a$, has density $\widehat{p}_t$ on $[a, b]$. If the fields and SDE extend to $t = 0, 1$ as specified below, it transports p_0 to p_1. The conclusion concerns the one-time marginals of the constructed probability path, not equality with the exact Brownian SB path law.

Matching equivalence and conditional regression. The conditional law of Z, given t and $X_t = x$, is

$$\mathbb{P}(Z \in \mathrm{d}z \mid t, X_t = x) = \frac{\widehat{p}_t(x \mid z)}{\widehat{p}_t(x)}\, \widehat{\pi}(\mathrm{d}z), \qquad \widehat{p}_t(x) > 0. \tag{A.38}$$

For any vector target U with conditional mean $m = \mathbb{E}[U \mid t, X_t = x]$, the positive-definite matrix $G(x)$ is fixed under this conditional expectation. Hence

$$\mathbb{E}[\|w - U\|^2_{g(x)} \mid t, X_t = x] = \|w - m\|^2_{g(x)} + \mathbb{E}[\|U - m\|^2_{g(x)} \mid t, X_t = x]. \tag{A.39}$$

The second term does not depend on w. The conditional mean of the flow target is $u_t^\circ(x)$ as defined in Theorem 4.4. For the score target,

$$\begin{aligned} \widehat{p}_t(x)\mathbb{E}[\nabla_g \log \widehat{p}_t(x \mid Z) \mid t, X_t = x] &= \int \widehat{p}_t(x \mid z) \nabla_g \log \widehat{p}_t(x \mid z)\, \widehat{\pi}(\mathrm{d}z) \\ &= \nabla_g \int \widehat{p}_t(x \mid z)\, \widehat{\pi}(\mathrm{d}z) = \nabla_g \widehat{p}_t(x). \end{aligned} \tag{A.40}$$

Thus its conditional mean is $\nabla_g \log \widehat{p}_t$. Under the additional representability assumption, (A.39) identifies the population minimizers as $v_{\theta^\star} = u_t^\circ$ and $s_{\theta^\star} = \nabla_g \log \widehat{p}_t$. Integrating the conditional continuity equation against $\widehat{\pi}$ also gives

$$\partial_t \widehat{p}_t = -\operatorname{div}_g\left(\int \widehat{p}_t(x \mid z) u_t^\circ(x \mid z)\, \widehat{\pi}(\mathrm{d}z)\right) = -\operatorname{div}_g(\widehat{p}_t v_{\theta^\star}). \tag{A.41}$$

Applying (A.39) to both targets and averaging over (t, X_t) gives $\mathcal{L}_{\text{SSR-FM}} = \mathcal{L}_{\text{marg}} + C$, where

$$\begin{aligned} C = \mathbb{E}\big[&\|u_t^\circ(X_t \mid Z) - u_t^\circ(X_t)\|^2_{g(X_t)} \\ &+ \lambda(t)^2 \|\nabla_g \log \widehat{p}_t(X_t \mid Z) - \nabla_g \log \widehat{p}_t(X_t)\|^2_{g(X_t)}\big]. \end{aligned}$$

This nonnegative constant is independent of θ. Differentiating the loss identity gives the gradient equality in (4.10), proving Theorem 4.4. These identities hold for every admissible θ and do not require the networks to attain the population optimum. The metric changes the squared norm but not the argument, because $G(X_t)$ is fixed when conditioning on (t, X_t). For the full training-time law, the same weighted integrability is required near $t = 0, 1$, where bridge targets may be singular.

SDE reconstruction. Denote the actual density of (5.1) at the population optimum by ρ_t. The reference terms c_g and B_G contribute $\varepsilon\Delta_g/2$, so

$$\partial_t \rho_t = -\operatorname{div}_g\left[\rho_t\left(v_{\theta^\star} + \frac{\varepsilon}{2} s_{\theta^\star}\right)\right] + \frac{\varepsilon}{2}\Delta_g \rho_t. \tag{A.42}$$

Substituting $\rho_t = \widehat{p}_t$ and $s_{\theta^\star} = \nabla_g \log \widehat{p}_t$ gives

$$\begin{aligned} &-\operatorname{div}_g(\widehat{p}_t v_{\theta^\star}) - \frac{\varepsilon}{2}\operatorname{div}_g(\widehat{p}_t \nabla_g \log \widehat{p}_t) + \frac{\varepsilon}{2}\Delta_g \widehat{p}_t \\ &\quad = -\operatorname{div}_g(\widehat{p}_t v_{\theta^\star}) = \partial_t \widehat{p}_t. \end{aligned} \tag{A.43}$$

Here $\widehat{p}_t \nabla_g \log \widehat{p}_t = \nabla_g \widehat{p}_t$ cancels the score and diffusion terms. With $\rho_a = \widehat{p}_a$, uniqueness of the initial-value problem gives $\rho_t = \widehat{p}_t$ on $[a, b]$.

Initial and final marginals. On the regular branch, $\gamma_t \to x_0$ and $\Sigma_t = tG(x_0)^{-1} + O(t^2)$ as $t \downarrow 0$; the analogous relation holds at x_1 as $t \uparrow 1$. The Gaussian conditionals therefore converge weakly to the corresponding point masses. Since $\widehat{\pi}$ has marginals p_0, p_1, their mixture converges to p_0, p_1. To identify the SDE initialized at p_0 on the full interval, the ideal fields must extend to the open interval with these limiting marginals, and the SDE and its weak Fokker–Planck problem must be well posed with initial law p_0. Training restricted to $[\delta, 1-\delta]$ alone does not establish this extension.

A.9 Large Deviations and Stochastic Fluctuations

In this section, we explain how large-deviation concentration and local Gaussian fluctuations describe the small-noise construction in Sections 4.1–4.2. The limit $\varepsilon \downarrow 0$ identifies the fluctuation scale; it does not prescribe setting $\varepsilon = 0$ in Algorithm 1.

Concentration and the action cost. For a precise path-space statement, consider a compact, connected, smooth Riemannian manifold without boundary. Write $Q_\varepsilon^{g,z}$ for the bridge of the Brownian reference with generator $\varepsilon\Delta_g/2$, conditioned on endpoints $z = (x_0, x_1)$. This family satisfies a large-deviation principle at speed $1/\varepsilon$, with good rate function

$$I_z(\omega) = \frac{1}{2}\int_0^1 \|\dot{\omega}_t\|^2_{g(\omega_t)}\,\mathrm{d}t - e(x_0, x_1) = S_g(\omega) - e(x_0, x_1), \tag{A.44}$$

for absolutely continuous paths with the prescribed endpoints and finite action, and $I_z = +\infty$ otherwise (Inahama, 2024, Theorem 2.1, specialized to the Riemannian case and our noise convention). If the minimizing geodesic γ is unique, it is the unique zero of I_z. Every closed set of paths separated from γ in the uniform topology then has strictly positive minimum rate and exponentially small bridge probability as $\varepsilon \downarrow 0$.

The corresponding logarithmic heat-kernel asymptotic gives $-\varepsilon \log k_\varepsilon^g(x_0, x_1) \to e(x_0, x_1)$ (Varadhan, 1967), explaining the leading action cost in Proposition 4.2. Large deviations alone do not

determine the factor $J^{-1/2}$: the volume correction $\frac{\varepsilon}{2}\log J$ follows from the finer heat-kernel expansion in Appendix A.4.

Fluctuations around the geodesic. Concentration on a fixed spatial scale is compatible with nontrivial fluctuations on the smaller $\sqrt{\varepsilon}$ scale. Under the unique, non-conjugate minimizing-geodesic assumptions of Appendix A.4, the fluctuation theorem of Bailleul et al. (2021) gives, for a fixed $t \in (0,1)$ and in the time-independent local coordinates used in Appendix A.6,

$$\begin{aligned} X_t^\varepsilon &\xrightarrow{\mathbb{P}} \gamma_t(z), \\ \frac{X_t^\varepsilon - \gamma_t(z)}{\sqrt{\varepsilon}} &\xrightarrow{\mathrm{d}} \mathcal{N}(0, \Sigma_t(z)), \qquad \Sigma_t(z) = H_t(z)^{-1}. \end{aligned} \tag{A.45}$$

Here X_t^ε is distributed according to the exact reference bridge marginal. The second limit requires the local second-order analysis in Appendix A.6, not merely the large-deviation principle. For each fixed positive ε, the sampler (4.5) retains the corresponding Gaussian spread, with covariance $\varepsilon\Sigma_t(z)$. Replacing it by $\gamma_t(z)$ alone keeps the leading location but removes this conditional spread. This concentration concerns a fixed endpoint pair; mixing over endpoint pairs need not collapse the population distribution to a point.

Positive noise in SSR-FM. Algorithm 1 uses a fixed positive ε in both the coupling and the conditional sampler. Small noise makes the local expansion informative while preserving its leading stochastic fluctuations. Reducing ε changes the Brownian reference and hence the target Schrödinger bridge; it is not only a numerical-accuracy adjustment, nor does the analysis require an annealing schedule toward zero. The role of randomness here is to represent conditional spread. Under suitable well-posedness conditions, a probability-flow ODE can also reproduce the resulting one-time marginals through the continuity equation in Appendix A.8. A quantitative finite-noise approximation bound requires estimates beyond this large-deviation interpretation.

B Additional Experimental Results

This appendix describes the datasets and comparison methods, followed by additional results in the same order as Sections 6.1–6.3.

B.1 Datasets

Table B1 summarizes the data sources used in Sections 6.1–6.3, their geometric representations, and the held-out times. The experiments cover simulated distributions on curved surfaces, observations with known spherical or periodic geometry, and single-cell populations whose metric is estimated from data.

Simulated data. The two tasks use a saddle-shaped surface with curled sides (the rolled saddle) and a helicoid, each supporting a spiral source and an S-shaped target in $\mathbb{R}^3$. Training receives 256 samples from each endpoint distribution. An independently computed finite-volume Brownian-SB solution on the surface patch provides the unobserved midpoint distribution. This numerical reference makes it possible to evaluate interpolation between the observed endpoints; it is not an exact bridge on the complete surface.

NOAA. The Global Drifter Program provides observations of drifting buoys used to study ocean surface circulation (Elipot et al., 2016; 2022). We use its quality-controlled hourly positions, accessed through NOAA's AOML ERDDAP service. Latitude and longitude are mapped to the unit sphere, giving a three-dimensional representation with a prescribed spherical metric.

IBTrACS. The International Best Track Archive for Climate Stewardship combines tropical-cyclone track records from multiple agencies (Knapp et al., 2010). We use the since-1980 subset of version 4r01 (Gahtan et al., 2024). Cyclone-center positions are represented on the unit sphere, with the six ocean-basin groups retained as separate conditions.

ALA2 and Chignolin. These datasets contain molecular-dynamics trajectories describing changes in molecular conformation. ALA2 uses alanine-dipeptide shooting trajectories from the SimTK release (Chodera et al., 2006), while Chignolin uses the released simulation coordinates (Wang et al., 2019).

The representations contain two and 18 dihedral angles, respectively. Mapping each angle to its cosine and sine gives $T^2 \subset \mathbb{R}^4$ and $T^{18} \subset \mathbb{R}^{36}$. These flat-torus metrics describe angular geometry, not the full molecular kinetic metric.

EMT. The epithelial–mesenchymal transition data contain gene-expression snapshots collected during a change in cell phenotype (Cook & Vanderhyden, 2020). The released representation used here contains 3,133 cells at four snapshot labels, with each cell represented by a 10-dimensional embedding.

EB. The embryoid-body data capture gene-expression changes during differentiation (Moon et al., 2019). The released dataset contains 16,819 cells at five snapshot labels. We use its PCA representations in 10, 20, 50, and 100 dimensions to evaluate interpolation across different representation sizes.

CITE-seq. We use the RNA component of the released CITE-seq data from the Open Problems Multimodal Single-Cell Integration benchmark (Burkhardt et al., 2022). The input contains 31,240 cells at four snapshot labels, represented in 50 dimensions. Protein measurements are not included in this comparison.

Data splits. Each analytic-metric dataset is represented at nine times, with indices 2 and 6 jointly withheld. Snapshots are independently permuted, so the models do not receive trajectory correspondences. For single-cell data, each interior snapshot is withheld in turn, and time inputs are normalized ordinal labels. Standardization, LAND estimation, and coupling construction use only the training cells of each fold. The released low-dimensional embeddings are reused across folds; their original dimensionality reduction is not refitted after withholding a snapshot.

Table B1: Datasets in the order of the main experiments. Counts are training support points per endpoint for the synthetic tasks, observations per reporting time for the analytic-metric tasks, and total cells for the single-cell tasks. Single-cell t_i are ordinal snapshot labels.

Dataset	Geometry / representation	Sample count	Held-out times
Rolled saddle	Surface in $\mathbb{R}^3$	256 / endpoint	$t = 0.5$
Helicoid	Surface in $\mathbb{R}^3$	256 / endpoint	$t = 0.5$
NOAA	$S^2 \subset \mathbb{R}^3$	3,837 / time	6, 18 days
IBTrACS	$S^2 \subset \mathbb{R}^3$	4,398 / time	12, 36 h
ALA2	$T^2 \subset \mathbb{R}^4$	5,367 / time	25, 75 ps
Chignolin	$T^{18} \subset \mathbb{R}^{36}$	3,720 / time	12.5, 37.5 ns
EMT	10D, estimated LAND	3,133 total	t_1, t_2
EB	10/20/50/100D, estimated LAND	16,819 total	t_1, t_2, t_3
CITE-seq	50D, estimated LAND	31,240 total	t_1, t_2

B.2 Baseline Methods and Evaluation Metrics

The comparison includes optimal-transport-based flow matching, geometry-aware flow matching, and simulation-based Schrödinger bridge methods.

OT-CFM. Optimal transport conditional flow matching learns a vector field from straight-line interpolations between OT-coupled samples (Tong et al., 2024a). We include it as an OT-based deterministic baseline for distribution interpolation.

SF²M. Simulation-free score and flow matching combines an entropic transport coupling with conditional flow and score regression (Tong et al., 2024b). We implement its original Brownian-bridge formulation for the present benchmarks. On $\mathcal{M} = \mathbb{R}^d$ with $G(x) = I$, SSR-FM has $J = 1$ and $c_g = 0$, giving

$$\begin{aligned} \widehat{c}_\varepsilon(x_0, x_1) &= \tfrac{1}{2}\|x_1 - x_0\|_2^2, & \gamma_t(z) &= (1-t)x_0 + tx_1, \\ \Sigma_t(z) &= t(1-t)I, & \widehat{p}_t(x \mid z) &= \mathcal{N}(x; \gamma_t(z), \varepsilon t(1-t)I). \end{aligned} \tag{B.1}$$

The heat-kernel and conditional Gaussian expressions are exact in this case. With the same Brownian variance ε and regression weights, the coupling, flow and score targets, and training objective reduce exactly to those of Brownian-reference SF²M.

Metric FM. Metric flow matching learns approximate geodesic interpolants by minimizing energy under a Riemannian metric, then trains a transport field along these paths (Kapuśniak et al., 2024). It provides a geometry-aware ODE baseline, particularly relevant when the metric is estimated from data. We use the official model and training objectives with precomputed OT couplings. On the simulated data, MFM receives the same prescribed metric as SSR-FM, avoiding an additional source of error from metric estimation in this comparison. On the analytic-metric real-data tasks, it uses its data-estimated LAND metric.

RFM. Riemannian flow matching constructs conditional vector fields that respect a given manifold geometry (Chen & Lipman, 2024). Our analytic-metric implementation uses intrinsic OT pairing; the simulated-surface and single-cell implementations use independent endpoint pairs. Its ODE evolution is deterministic given an initial sample. With the same metric, endpoint coupling, and constant-speed geodesics, removing conditional fluctuations from SSR-FM gives $X_t = \gamma_t(z)$ and $u_t^\circ(X_t \mid z) = \dot{\gamma}_t(z)$; omitting score matching then recovers the geodesic RFM objective for that coupling. SSR-FM extends this deterministic construction by deriving the coupling and geometry-dependent conditional covariance from a common Riemannian Brownian reference through small-noise approximations. Joint flow and score matching then enables stochastic interpolation with the prescribed diffusion geometry while retaining simulation-free training.

RDSB. Riemannian diffusion Schrödinger bridges learn forward and backward diffusions using simulated paths (Thornton et al., 2022). This baseline addresses manifold-SB learning through iterative simulation, providing a comparison with the simulation-free training strategy. We reproduce RDSB following the algorithm in the original paper. We fit independent forward and backward fields for each pair of neighbouring training snapshots, using 10 simulation steps, two IPF rounds, and approximately 50,000 total updates per complete setting, shared across intervals and directions. We schedule five training seeds (0–4) for each setting and select the trained model for each direction using fixed validation simulations based only on training snapshots; the helicoid failure is reported in Appendix B.3.

Evaluation. We evaluate distribution matching using the 1-Wasserstein distance W_1 and, for datasets with a known manifold, geometric consistency using the mean distance to that manifold. At each evaluation time t, let $\widehat{p}_t = n^{-1}\sum_{i=1}^n \delta_{\widehat{x}_i}$ and $p_t^{\mathrm{data}} = m^{-1}\sum_{j=1}^m \delta_{y_j}$ denote the empirical predicted and observed distributions. Their distributional error is

$$W_1(\widehat{p}_t, p_t^{\mathrm{data}}) = \min_{P\in\Pi(a,b)} \sum_{i=1}^n \sum_{j=1}^m P_{ij}\|\widehat{x}_i - y_j\|_2, \qquad a_i = \frac{1}{n}, \quad b_j = \frac{1}{m}, \tag{B.2}$$

where $\Pi(a,b) = \{P \geq 0 : P\mathbf{1}_m = a,\ P^\top \mathbf{1}_n = b\}$. Each sample has uniform mass within its distribution, and the optimization finds a coupling without requiring paired trajectories. We solve this unregularized OT problem on the full predicted and observed populations for the real-data experiments. In the simulated-data experiments, we generate 2,560 trajectories per trained model and report the mean and sample standard deviation of W_1 over five evaluations, each using 512 generated samples and 512 reference samples. We use the Hungarian algorithm to solve each equal-size assignment problem, with the same reference samples across methods.

For NOAA and IBTrACS, the Euclidean distance between unit-sphere coordinates is multiplied by the Earth radius $R = 6371.0088$ km, so the reported W_1 uses chord distances in kilometers. For ALA2 and Chignolin, distances are computed in the full cosine–sine representation. Single-cell distances use the released embedding after standardization fitted on training snapshots. Where a dataset has multiple conditions, OT is solved separately within each condition before averaging the errors.

The empirical mean distance to the manifold is

$$D_{\mathcal{M}}(\widehat{p}_t) = \frac{1}{n}\sum_{i=1}^n \mathrm{dist}_2(\widehat{x}_i, \mathcal{M}), \qquad \mathrm{dist}_2(x,\mathcal{M}) = \inf_{y\in\mathcal{M}} \|x - y\|_2. \tag{B.3}$$

For the unit sphere and the product of d unit circles, respectively,

$$\mathrm{dist}_2(x, S^2) = \big|\|x\|_2 - 1\big|, \qquad \mathrm{dist}_2(x, T^d) = \left[\sum_{\ell=1}^{d} \big(\|x^{(\ell)}\|_2 - 1\big)^2\right]^{1/2}, \tag{B.4}$$

where $x^{(\ell)} \in \mathbb{R}^2$ is the coordinate pair for the ℓth angle. Spherical distances are again multiplied by R; the torus distance is not divided by $\sqrt{d}$. For the simulated data, the closest-point distance is computed numerically from the surface parameterization. Lower values indicate better preservation of the prescribed geometry. Predictions at all times are obtained by evolving the initial population continuously, without restarting from the observed intermediate snapshots.

B.3 Synthetic Interpolation on Curved Surfaces

Manifold geometry. We consider two surfaces embedded in $\mathbb{R}^3$, a rolled saddle and a helicoid, parameterized by

$$F_{\mathrm{sad}}(u, v) = \begin{pmatrix} u \\ \alpha v\big[1 - 0.6(v/1.25)^2\big] \\ 0.3u^2 - 0.06u - 0.7v^2 \end{pmatrix}, \qquad F_{\mathrm{hel}}(r, \varphi) = \begin{pmatrix} r\cos\varphi \\ r\sin\varphi \\ 0.5\varphi \end{pmatrix}, \tag{B.5}$$

where $\alpha \simeq 2.01246$. The cubic transverse coordinate curls the sides of the saddle. On the helicoid, r is a signed radius and φ is an unwrapped angle, so successive turns have different heights. Both tasks use the metric induced by their embedding:

$$G_{\mathrm{sad}} = (DF_{\mathrm{sad}})^\top DF_{\mathrm{sad}}, \qquad G_{\mathrm{hel}}(r, \varphi) = \begin{pmatrix} 1 & 0 \\ 0 & r^2 + 0.25 \end{pmatrix}. \tag{B.6}$$

The Brownian reference has generator $\varepsilon\Delta_g/2$, with $\varepsilon = 0.1$. Figure 1 shows the rolled-saddle experiment; Figure B1 provides the complementary helicoid visualization.

Data and training setting. Each surface supports a noisy spiral source distribution p_0 and an S-shaped target distribution p_1. All methods receive the same 256 independently sampled observations from each endpoint, with uniform empirical weights and no trajectory correspondences. Only $t = 0$ and $t = 1$ are observed. An independent finite-volume solution of the Brownian Schrödinger bridge provides the numerical reference at $t = 0.25, 0.5, 0.75$; these reference distributions are used only for evaluation, not for training or model selection.

The flow-matching fields are trained for 20,000 updates with batch size 512, sampling endpoint pairs with replacement from the fixed training observations. Training times are drawn uniformly from $[0.02, 0.98]$. Flow and score networks have three hidden layers of width 128. Field training uses AdamW with learning rate 10^{-3}, weight decay 10^{-5}, gradient clipping at 10, and an exponential moving average with decay 0.999. We use training seed 0 and the final trained model.

OT-CFM and SF²M are trained in the ambient space $\mathbb{R}^3$ with exact squared-distance OT and Brownian-reference Sinkhorn couplings, respectively. Metric FM uses the official two-stage GeoPath–flow implementation in surface coordinates with exact OT pairing and the prescribed metric; its GeoPath network receives 20,000 additional Adam updates at learning rate 10^{-3}, and its conditional noise parameter is $\sigma = 0.1$. RFM uses independent endpoint pairs and numerically computed geodesics under the same metric. For SSR-FM, we precompute action-minimizing paths and full action Hessians for all 256^2 endpoint pairs using 32 time intervals. SF²M and SSR-FM both use $\varepsilon = 0.1$. RDSB uses the pairwise IPF recipe in Appendix B.2, with 50,000 updates per seed divided equally among four directional stages.

Sampling. OT-CFM, Metric FM, and RFM use ODE sampling with Heun's method; SF²M, RDSB, and SSR-FM use SDE sampling with Euler–Maruyama. All samplers use 640 steps over $[0, 1]$ and the same 2,560 initial evaluation particles, separate from the training observations. We evaluate the saved trajectories at $t = 0.25, 0.5, 0.75$, reporting five evaluations of 512 generated and 512 reference samples at each time, as described in Appendix B.2.

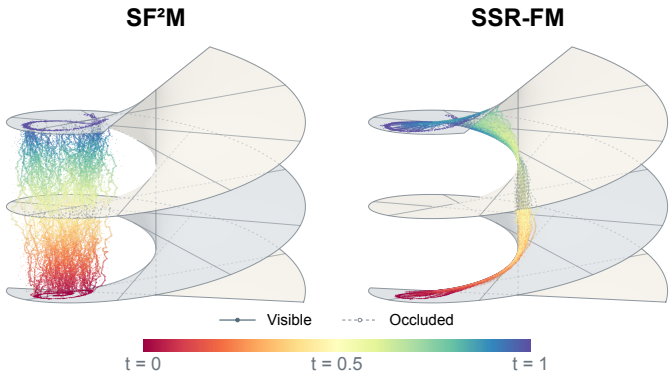


Figure B1: Endpoint-only interpolation on the helicoid. SF²M (left) produces off-surface intermediate states, whereas SSR-FM (right) follows the surface. Colors indicate time. Lines connect saved reporting nodes and are not exact continuous Brownian trajectories. Endpoint clouds depict the prescribed source and target distributions, not the learned endpoint fit.

Results. Tables B2 and B3 report W_1 and $D_{\mathcal{M}}$ at three unobserved times on each surface. SSR-FM achieves the lowest W_1 in all six dataset–time settings, while retaining near-zero distance to the surface. For RDSB, we first average the five evaluations within each trained model, then report the mean and sample standard deviation across training seeds. For the other methods, the standard deviations describe evaluation-sample variation for a fixed model trained with seed 0. Rankings use unrounded W_1 means; manifold-distance ties use the displayed precision.

Table B2: Rolled-saddle interpolation at three unobserved times. Mean ± sample SD. RDSB: five training seeds (0–4). Other methods: five evaluations of one model trained with seed 0 (512 samples per distribution per evaluation). Lower is better. Bold and underlining mark the best and second-best means; tied zeros are bold only.

Method	$t = 0.25$		$t = 0.5$		$t = 0.75$	
	W_1	$D_{\mathcal{M}}$	W_1	$D_{\mathcal{M}}$	W_1	$D_{\mathcal{M}}$
OT-CFM	0.500 ±0.013	0.421 ±0.002	0.704 ±0.009	0.680 ±0.004	0.519 ±0.015	0.477 ±0.003
SF²M	0.499 ±0.011	0.427 ±0.008	0.698 ±0.015	0.679 ±0.016	0.521 ±0.021	0.478 ±0.008
Metric FM	<u>0.124 ±0.022</u>	**0.000 ±0.000**	<u>0.136 ±0.026</u>	**0.000 ±0.000**	0.148 ±0.033	**0.000 ±0.000**
RFM	0.169 ±0.003	**0.000 ±0.000**	0.173 ±0.013	**0.000 ±0.000**	<u>0.143 ±0.028</u>	**0.000 ±0.000**
RDSB	0.539 ±0.142	**0.000 ±0.000**	0.778 ±0.281	**0.000 ±0.000**	0.897 ±0.346	0.001 ±0.002
SSR-FM	**0.097 ±0.019**	**0.000 ±0.000**	**0.109 ±0.022**	**0.000 ±0.000**	**0.118 ±0.022**	**0.000 ±0.000**

On the rolled saddle, SSR-FM reduces W_1 relative to the strongest baseline by 22.2%, 19.7%, and 17.7% at $t = 0.25, 0.5, 0.75$, respectively. Metric FM and RFM also remain on the surface, but their larger W_1 values show that geometric adherence alone is insufficient to recover the intermediate distributions. OT-CFM and SF²M have both larger distributional errors and substantial off-surface displacement at all three times.

On the helicoid, SSR-FM improves on RFM, the strongest baseline at each time, by 17.7%, 34.7%, and 12.1%, respectively. The time-resolved geometric errors also clarify the visualization in Figure B1: OT-CFM has a small manifold distance of 0.0076 at the midpoint, yet its distances at $t = 0.25$ and $t = 0.75$ are 0.7514 and 0.7513. Proximity to the surface at one instant therefore does not imply geometric adherence throughout the interpolation. SSR-FM maintains near-zero manifold distance at all three reporting times and more accurately matches the reference distributions. These comparisons concern the complete methods, with different couplings and conditional paths, rather than isolating the effect of stochasticity.

RDSB failure on the helicoid. We attempted five training seeds (0–4), but seed 4 terminated with a non-finite gradient norm; its last saved model was obtained after 25,000 of 50,000 updates. The

Table B3: Helicoid interpolation at three unobserved times. Mean ± sample SD. RDSB: four successful training seeds (0–3) out of five attempted; seed 4 failed. Other methods: five evaluations of one model trained with seed 0 (512 samples per distribution per evaluation). Lower is better. Bold and underlining mark the best and second-best means; tied zeros are bold only.

Method	$t = 0.25$		$t = 0.5$		$t = 0.75$	
	W_1	$D_{\mathcal{M}}$	W_1	$D_{\mathcal{M}}$	W_1	$D_{\mathcal{M}}$
OT-CFM	1.598 ±0.011	0.751 ±0.000	2.002 ±0.012	0.008 ±0.000	1.636 ±0.027	0.751 ±0.001
SF²M	1.599 ±0.014	0.649 ±0.004	1.993 ±0.016	0.129 ±0.004	1.637 ±0.025	0.647 ±0.002
Metric FM	0.181 ±0.011	**0.000 ±0.000**	0.111 ±0.006	**0.000 ±0.000**	0.209 ±0.019	**0.000 ±0.000**
RFM	<u>0.146 ±0.009</u>	**0.000 ±0.000**	<u>0.091 ±0.013</u>	**0.000 ±0.000**	<u>0.209 ±0.017</u>	**0.000 ±0.000**
RDSB	1.009 ±0.188	0.001 ±0.002	1.541 ±0.199	0.007 ±0.011	2.231 ±0.343	0.005 ±0.011
SSR-FM	**0.120 ±0.012**	**0.000 ±0.000**	**0.060 ±0.015**	**0.000 ±0.000**	**0.184 ±0.028**	**0.000 ±0.000**

helicoid RDSB entries in Tables 1 and B3 therefore summarize the four completed runs (seeds 0–3), without replacing the failed seed or imputing its result. Their sample standard deviations describe variability among successful runs only.

B.4 Experiments with Analytic Metrics

Overall results. Table B4 adds sample standard deviations to the dataset-level comparison in Table 2. We first average the two held-out-time errors within each training seed, then compute the mean and sample standard deviation across seeds. The following subsections give the settings and time-resolved results.

Table B4: Dataset-level interpolation results corresponding to Table 2. Mean ± sample SD over five training seeds (0–4), except Metric FM on NOAA and IBTrACS (four retained runs). Geographic errors are in km; molecular errors use cosine–sine embedding units. Lower is better.

Geographic datasets

Method	NOAA (S^2)		IBTrACS (S^2)	
	W_1	$D_{\mathcal{M}}$	W_1	$D_{\mathcal{M}}$
OT-CFM	108.418 ±0.186	5.192 ±0.249	126.555 ±0.118	9.917 ±0.331
SF²M (ODE)	102.102 ±0.650	6.744 ±0.834	<u>119.842</u> ±0.240	8.273 ±0.446
SF²M (SDE)	102.323 ±0.648	7.801 ±0.695	125.202 ±0.223	24.965 ±0.159
Metric FM	113.207 ±1.319	12.279 ±3.193	129.900 ±0.988	14.193 ±2.490
RFM	<u>101.273</u> ±0.103	**0.000** ±0.000	122.775 ±0.250	0.022 ±0.000
RDSB	127.715 ±0.055	<u>0.003</u> ±0.000	282.288 ±0.172	**0.010** ±0.000
SSR-FM (ODE)	**101.228** ±0.588	0.004 ±0.000	**119.722** ±0.425	0.020 ±0.000
SSR-FM (SDE)	101.284 ±0.582	0.006 ±0.000	120.159 ±0.415	<u>0.011</u> ±0.000

Molecular datasets

Method	ALA2 (T^2)		Chignolin (T^{18})	
	W_1	$D_{\mathcal{M}}$	W_1	$D_{\mathcal{M}}$
OT-CFM	0.319 ±0.003	0.138 ±0.004	**1.951** ±0.010	0.618 ±0.014
SF²M (ODE)	0.398 ±0.037	0.077 ±0.005	2.099 ±0.026	0.765 ±0.023
SF²M (SDE)	0.351 ±0.034	0.074 ±0.003	2.105 ±0.024	0.763 ±0.014
Metric FM	0.629 ±0.195	0.292 ±0.122	2.097 ±0.130	0.765 ±0.175
RFM	0.316 ±0.002	**0.000** ±0.000	2.354 ±0.017	**0.000** ±0.000
RDSB	1.262 ±0.000	**0.000** ±0.000	2.469 ±0.001	**0.000** ±0.000
SSR-FM (ODE)	<u>0.314</u> ±0.004	**0.000** ±0.000	2.129 ±0.013	**0.000** ±0.000
SSR-FM (SDE)	**0.308** ±0.004	**0.000** ±0.000	<u>1.985</u> ±0.007	**0.000** ±0.000

B.4.1 GEOGRAPHIC DATA: NOAA AND IBTRACS

Manifold geometry. NOAA drifters and IBTrACS cyclone centers are represented on the unit sphere with its induced metric,

$$S^2 = \{x \in \mathbb{R}^3 : \|x\|_2 = 1\}, \qquad g_x(v, w) = v^\top w, \quad v, w \in T_x S^2. \tag{B.7}$$

As in Appendix B.2, both distributional and manifold errors are reported in kilometers.

Data and training setting. Each task contains nine reporting times, with the third and seventh snapshots jointly withheld and the other seven supplied as unpaired populations. The hidden times are 6 and 18 days for NOAA and 12 and 36 hours for IBTrACS. Each snapshot contains 3,837 and 4,398 observations, respectively; IBTrACS retains six basin conditions. Snapshots are independently permuted within each condition, so no method receives trajectory correspondences. Training uses physical times normalized to $[0, 1]$ and only the seven training populations.

The flow-matching fields use the network and optimizer settings in Appendix B.3: 20,000 updates with batch size 512 and learning rate 10^{-3}. We train with five random seeds (0–4).

OT-CFM and SF²M operate in the ambient representation; OT-CFM uses conditional noise $\sigma = 0.1$. Metric FM follows the official two-stage implementation with a LAND metric estimated from training observations; its GeoPath stage uses 20,000 additional updates at learning rate 0.02, and $\sigma = 0.1$. RFM uses the prescribed spherical geometry and intrinsic OT pairing. SSR-FM constructs full-support Sinkhorn couplings between neighbouring training populations using the spherical Brownian reference. SF²M and SSR-FM both use the data-scaled variance rate $\varepsilon = \beta s_{\mathrm{data}}$ with $\beta = 0.005$ on NOAA and IBTrACS. The shared scale s_{data} is the median of squared intrinsic OT-pair distances divided by twice the corresponding normalized time interval, estimated from training snapshots. RDSB follows the original alternating forward/backward IPF algorithm independently on the six training intervals, using 10 simulation steps and two IPF rounds, with 49,992 training updates in total for each of five seeds (0–4).

Sampling. All methods propagate the full initial population continuously over the observation horizon using 640 integration steps, without resetting particles at training times. SF²M and SSR-FM report flow-only ODE sampling and SDE sampling from the same trained model separately. Stochastic sampling is repeated five times per seed for NOAA and 16 for IBTrACS. We average evaluation errors across repetitions within each seed, then report the mean and sample standard deviation across training seeds, with the Metric FM exclusions described below.

Table B5: NOAA (km) interpolation at the two jointly held-out times. Mean ± sample SD over five training seeds (0–4), except Metric FM (four retained runs). Lower is better. Bold and underlining mark the best and second-best means; tied zeros are bold only.

Method	$t = 6$ days		$t = 18$ days	
	W_1	$D_{\mathcal{M}}$	W_1	$D_{\mathcal{M}}$
OT-CFM	81.070 ±0.103	2.838 ±0.118	135.766 ±0.270	7.547 ±0.403
SF²M (ODE)	77.359 ±0.371	3.686 ±0.403	126.846 ±0.932	9.802 ±1.268
SF²M (SDE)	77.542 ±0.371	4.611 ±0.315	127.104 ±0.932	10.992 ±1.080
Metric FM	82.816 ±0.588	6.235 ±1.463	143.598 ±2.053	18.324 ±4.932
RFM	**75.141 ±0.297**	**0.000 ±0.000**	127.405 ±0.366	**0.000 ±0.000**
RDSB	88.968 ±0.042	0.002 ±0.000	166.462 ±0.083	0.005 ±0.000
SSR-FM (ODE)	76.890 ±0.413	0.002 ±0.000	**125.565 ±0.775**	0.006 ±0.000
SSR-FM (SDE)	76.951 ±0.411	0.003 ±0.000	125.618 ±0.765	0.009 ±0.000

Results. Tables B5 and B6 report W_1 and $D_{\mathcal{M}}$ at each hidden time. Averaged over the two times, SSR-FM (ODE) has the lowest W_1 on both datasets. Bold and underlining follow the ranking convention in Appendix B.3, including the standard deviations. For Metric FM, all five seeds were trained, but post-hoc inspection identified unstable GeoPath optimization in NOAA seed 0 and IBTrACS seed 3. Their normalized GeoPath energies ended at 121.7 and 21.1, respectively, and the subsequent flow losses remained elevated. Both runs completed with finite values. We exclude these

Table B6: IBTrACS (km) interpolation at the two jointly held-out times. Mean ± sample SD over five training seeds (0–4), except Metric FM (four retained runs). Lower is better. Bold and underlining mark the best and second-best means; tied zeros are bold only.

Method	$t = 12$ h		$t = 36$ h	
	W_1	$D_{\mathcal{M}}$	W_1	$D_{\mathcal{M}}$
OT-CFM	96.148 ±0.072	4.999 ±0.220	156.961 ±0.175	14.836 ±0.468
SF²M (ODE)	<u>93.950 ±0.134</u>	4.552 ±0.217	<u>145.734 ±0.356</u>	11.994 ±0.682
SF²M (SDE)	97.992 ±0.131	17.693 ±0.060	152.411 ±0.321	32.237 ±0.259
Metric FM	97.875 ±0.736	8.000 ±2.419	161.924 ±1.349	20.386 ±2.587
RFM	94.977 ±0.105	0.012 ±0.000	150.574 ±0.412	0.033 ±0.000
RDSB	155.505 ±0.115	**0.005 ±0.000**	409.070 ±0.244	**0.015 ±0.000**
SSR-FM (ODE)	**93.896 ±0.131**	<u>0.010 ±0.000</u>	**145.549 ±0.721**	0.029 ±0.000
SSR-FM (SDE)	94.512 ±0.131	**0.005 ±0.000**	145.806 ±0.701	<u>0.016 ±0.000</u>

two runs from the reported Metric FM means and sample standard deviations, retaining seeds 1–4 on NOAA and 0, 1, 2, 4 on IBTrACS. Including all five runs gives two-time mean $(W_1, D_{\mathcal{M}})$ values of $(133.619, 24.196)$ km and $(151.933, 30.042)$ km, respectively; the first- and second-ranked W_1 entries are unchanged. These runs indicate seed-sensitive training under the current setting.

On NOAA, SSR-FM (ODE) is second at 6 days and first at 18 days, giving the lowest two-time average of 101.228 km, closely followed by RFM at 101.273 km.

On IBTrACS, SSR-FM (ODE) and SF²M (ODE) are first and second, respectively, at both hidden times. SSR-FM (ODE) reduces the two-time mean W_1 from SF²M's 119.842 to 119.722 km and manifold distance from 8.273 to 0.020 km. With matched $\beta = 0.005$, SSR-FM (SDE) reduces W_1 from SF²M's 125.202 to 120.159 km and manifold distance from 24.965 to 0.011 km. Across the geographic settings, both SSR-FM samplers retain manifold distances of 0.002–0.029 km (approximately 2–29 m). The advantage is the combination of accurate distribution matching and geometric adherence, rather than the smallest manifold distance in every column.

B.4.2 Molecular Data: ALA2 and Chignolin

Manifold geometry. ALA2 and Chignolin are represented on flat tori with $d = 2$ and $d = 18$ dihedral angles, respectively. Their cosine–sine embedding and metric are

$$\begin{aligned} F_d(\theta) &= (\cos\theta_1, \sin\theta_1, \ldots, \cos\theta_d, \sin\theta_d), \\ G_d(\theta) &= (DF_d(\theta))^\top DF_d(\theta) = I_d, \qquad \theta \in (\mathbb{R}/2\pi\mathbb{Z})^d. \end{aligned} \tag{B.8}$$

This geometry accounts for angular periodicity. Both W_1 and $D_{\mathcal{M}}$ use the full $2d$-dimensional cosine–sine representation.

Data and training setting. The two tasks use the same nine-snapshot split as the geographic experiments. ALA2 contains 5,367 observations per snapshot across six initial-state conditions, with 25 and 75 ps withheld. Chignolin contains 3,720 observations per snapshot, with 12.5 and 37.5 ns withheld. All snapshots are independently permuted within each condition, and only the seven training populations are supplied during training.

We use the training budgets, network and optimizer settings, and method-specific constructions in Appendix B.4.1, replacing spherical geometry by the periodic torus geometry for the intrinsic methods. For SSR-FM, the periodic Brownian reference uses $\varepsilon = \beta s_{\mathrm{data}}$, where s_{data} is estimated from OT-matched samples from neighbouring training snapshots. We divide each squared periodic distance by the torus dimension and the corresponding time interval, then take the median. We use $\beta = 0.005$ on ALA2 and $\beta = 0.1$ on Chignolin. SF²M uses $\varepsilon = 0.01$ on both. Metric FM estimates LAND from the training observations. RDSB uses the same pairwise IPF setting with 49,992 updates per seed. All methods report five training seeds (0–4).

Sampling. Sampling again uses 640 steps with no resets at training times, with 16 stochastic repetitions per seed.

ALA2 results. SSR-FM (SDE) gives the lowest W_1 at both hidden times, with a two-time mean of 0.308 versus 0.316 for the strongest baseline, RFM. Table B7 reports the full comparison. Its manifold distance rounds to 0.000 at both times, whereas OT-CFM and SF²M have nonzero geometric errors.

Table B7: ALA2 interpolation at the two jointly held-out times. Mean ± sample SD over five training seeds (0–4). Lower is better. Bold and underlining mark the best and second-best means; tied zeros are bold only.

Method	$t = 25$ ps		$t = 75$ ps	
	W_1	$D_{\mathcal{M}}$	W_1	$D_{\mathcal{M}}$
OT-CFM	$\underline{0.364 \pm 0.004}$	0.183 ± 0.007	0.275 ± 0.007	0.092 ± 0.005
SF²M (ODE)	0.394 ± 0.059	0.094 ± 0.012	0.401 ± 0.019	0.060 ± 0.004
SF²M (SDE)	0.381 ± 0.051	0.099 ± 0.004	0.321 ± 0.020	0.049 ± 0.003
Metric FM	0.589 ± 0.136	0.289 ± 0.086	0.668 ± 0.268	0.295 ± 0.174
RFM	0.369 ± 0.006	$\mathbf{0.000 \pm 0.000}$	$\underline{0.264 \pm 0.004}$	$\mathbf{0.000 \pm 0.000}$
RDSB	1.098 ± 0.000	$\mathbf{0.000 \pm 0.000}$	1.425 ± 0.000	$\mathbf{0.000 \pm 0.000}$
SSR-FM (ODE)	0.365 ± 0.011	$\mathbf{0.000 \pm 0.000}$	0.264 ± 0.006	$\mathbf{0.000 \pm 0.000}$
SSR-FM (SDE)	$\mathbf{0.361 \pm 0.011}$	$\mathbf{0.000 \pm 0.000}$	$\mathbf{0.256 \pm 0.005}$	$\mathbf{0.000 \pm 0.000}$

Chignolin results. SSR-FM (SDE) is best at 37.5 ns and second when averaged over both hidden times: its mean W_1 of 1.985 is within 1.8% of OT-CFM's 1.951 (Table B8). Its manifold distance rounds to 0.000 at both times, compared with 0.504 and 0.732 for OT-CFM. At 12.5 ns, however, OT-CFM is best and SF²M (ODE) is second; SSR-FM is not among the two lowest errors at this time.

Table B8: Chignolin interpolation at the two jointly held-out times. Mean ± sample SD over five training seeds (0–4). Lower is better. Bold and underlining mark the best and second-best means; tied zeros are bold only.

Method	$t = 12.5$ ns		$t = 37.5$ ns	
	W_1	$D_{\mathcal{M}}$	W_1	$D_{\mathcal{M}}$
OT-CFM	$\mathbf{1.871 \pm 0.006}$	0.504 ± 0.011	$\underline{2.031 \pm 0.014}$	0.732 ± 0.018
SF²M (ODE)	$\underline{1.926 \pm 0.013}$	0.572 ± 0.014	2.272 ± 0.040	0.958 ± 0.033
SF²M (SDE)	1.935 ± 0.013	0.592 ± 0.010	2.274 ± 0.036	0.934 ± 0.020
Metric FM	1.956 ± 0.040	0.594 ± 0.079	2.237 ± 0.222	0.935 ± 0.275
RFM	2.132 ± 0.010	$\mathbf{0.000 \pm 0.000}$	2.576 ± 0.026	$\mathbf{0.000 \pm 0.000}$
RDSB	2.375 ± 0.001	$\mathbf{0.000 \pm 0.000}$	2.563 ± 0.002	$\mathbf{0.000 \pm 0.000}$
SSR-FM (ODE)	2.037 ± 0.009	$\mathbf{0.000 \pm 0.000}$	2.222 ± 0.020	$\mathbf{0.000 \pm 0.000}$
SSR-FM (SDE)	1.987 ± 0.006	$\mathbf{0.000 \pm 0.000}$	$\mathbf{1.984 \pm 0.009}$	$\mathbf{0.000 \pm 0.000}$

RFM and RDSB also preserve the torus geometry, but have larger W_1 than SSR-FM (SDE) on both molecular datasets. Taken together, the results support a favorable combination of distribution accuracy and geometric adherence: SSR-FM is first or second in dataset-level W_1 across the four analytic-metric tasks, but does not lead at every individual time.

B.5 Experiments with Data-Estimated Metrics

Data and metric. These experiments test interpolation when the geometry must be estimated from observed populations rather than supplied analytically. We use the released 10-dimensional EMT embedding, 50-dimensional CITE-seq RNA embedding, and EB principal components in 10, 20, 50, and 100 dimensions. EMT and CITE-seq each have four snapshots, whereas EB has five. Writing their normalized observation times as $t_k = k/K$, with $K = 3$ or 4, we withhold each interior

snapshot in turn and train on all cells at the remaining times. This gives 16 dataset–time settings, with no paired cell trajectories. Coordinate standardization and metric construction are fitted separately to the training cells of each fold; the released embeddings themselves are not recomputed.

The geometry-aware methods share a diagonal LAND metric (Arvanitidis et al., 2016),

$$G(x) = \mathrm{diag}\big(g_1(x), \ldots, g_d(x)\big),$$
$$g_j(x) = c\left[\rho + \sum_{a\in\mathcal{A}} \exp\left(-\frac{\|x-a\|_2^2}{2\gamma^2}\right)(x_j - a_j)^2\right]^{-\alpha}. \tag{B.9}$$

Here $\mathcal{A}$ contains 128 cells sampled without replacement from each training snapshot. These reference cells are the centers of the Gaussian distance weights. The bandwidth γ controls their spatial range: smaller values give more local weights, while larger values average over a wider neighborhood. We set γ to half the median radius from 26-neighbor queries within each reference cell's snapshot, and use $\rho = 10^{-4}$ and $\alpha = 0.3$. The factor c normalizes the average diagonal cometric on training cells; $G(x)$ varies with the cell state x.

Training setting. The flow-matching fields use 20,000 updates with batch size 512, AdamW learning rate 10^{-3}, weight decay 10^{-5}, gradient clipping at 10, and an exponential moving average with decay 0.999. We use smaller networks for low-dimensional data (EMT and EB10D/20D) and larger networks for high-dimensional data (CITE-seq and EB50D/100D). Both have three hidden layers, with widths of 128 and 512, respectively. We train one global-time field across training intervals and sample intervals proportionally to their durations. We report the final results from five training runs.

For training snapshots, OT-CFM and Metric FM use minibatch OT, SF²M uses Brownian-reference Sinkhorn couplings, and RFM samples independent endpoint pairs. Metric FM additionally trains its official GeoPath network for 20,000 Adam updates at learning rate 0.02, with conditional noise parameter $\sigma = 0.1$. SF²M and SSR-FM share the noise calibration based on training data, with multiplier 0.005.

SSR-FM computation. Within each minibatch, SSR-FM retains the nonzero pairs from an exact OT plan together with the top 32 entries of each row and column of an entropic proposal. For these candidate pairs, we minimize the Riemannian Brownian action under the shared LAND metric using four path segments. The inverse metric accumulated along each path gives a local quadratic covariance and the corresponding cost prefactor, avoiding formation of the full action Hessian. Sparse Sinkhorn then computes the endpoint coupling. Training samples endpoint pairs from this coupling and directly draws intermediate Gaussian states for flow and score regression. Appendix C gives the computational procedure, formulas, and controlled comparison.

Sampling. We propagate every cell in the initial snapshot continuously over $[0, 1]$, without restarting from intermediate training populations: 577 initial cells for EMT, 2,381 for EB, and 7,476 for CITE-seq. ODE sampling uses explicit Euler integration and SDE sampling uses Euler–Maruyama, with a nominal 640-step grid augmented by the observation times. SF²M and SSR-FM additionally report flow-only ODE sampling from the same trained model. We compute W_1 by exact balanced OT between the full predicted and held-out populations in the standardized coordinates. Since these datasets do not provide a known embedded manifold, we do not report $D_{\mathcal{M}}$. All methods report means over five training seeds (0–4), with sample standard deviations. RDSB uses 49,992–50,000 total updates per setting, shared across training intervals and directions.

Overall results. Table B9 adds sample standard deviations to the six-task comparison in Table 3, using the same held-out snapshots as the main text. We then report every interior-time holdout separately.

EMT and CITE-seq results. Table B10 shows that SSR-FM (ODE) gives the lowest W_1 at EMT t_2 and CITE-seq t_1. At EMT t_2, its SDE variant is second; at CITE-seq t_2, RFM is best and RDSB is second with $W_1 = 6.019$. SSR-FM (SDE) is third with 6.064, improving on SF²M (SDE) at 6.244. The advantage is not uniform across times: at EMT t_1, SF²M (ODE) and OT-CFM are first and second, respectively.

EB in 10D and 20D. Table B11 shows that SSR-FM (SDE) is best at t_2 in both representations. For EB20D, SSR-FM (ODE) is best at t_1, and its SDE variant is second at t_1 and t_3. On EB10D, SSR-FM (ODE) is second at t_1, but both SSR-FM samplers have higher errors than their SF²M counterparts

Table B9: Single-cell interpolation results corresponding to Table 3, with mean ± sample SD over five training seeds (0–4). The held-out snapshot is t_2 for EMT and EB, and t_1 for CITE-seq. Lower W_1 is better.

Method	EMT (10D)	EB (10D)	EB (20D)	EB (50D)	EB (100D)	CITE-seq (50D)
OT-CFM	2.1214 ±0.0160	1.5155 ±0.0260	3.1147 ±0.0192	7.1398 ±0.0263	11.5376 ±0.0410	6.0854 ±0.0028
SF²M (ODE)	2.1136 ±0.0079	1.5181 ±0.0067	3.1069 ±0.0060	6.6484 ±0.0174	10.6704 ±0.0083	6.0717 ±0.0026
SF²M (SDE)	2.1125 ±0.0083	1.5179 ±0.0042	3.0958 ±0.0077	6.4114 ±0.0083	10.4378 ±0.0045	6.1131 ±0.0014
Metric FM	2.1073 ±0.0064	1.5580 ±0.0066	3.6554 ±1.1605	6.7708 ±0.0234	11.2393 ±0.0217	6.0527 ±0.0039
RFM	2.1682 ±0.0050	1.6987 ±0.0102	3.1890 ±0.0162	**6.1324** ±0.0081	10.4497 ±0.0690	6.0359 ±0.0026
RDSB	2.2699 ±0.0139	2.9997 ±0.0100	3.6453 ±0.0220	6.5464 ±0.0209	10.9956 ±0.0106	6.3635 ±0.0052
SSR-FM (ODE)	**2.0977** ±0.0082	1.5174 ±0.0016	3.0991 ±0.0058	6.5079 ±0.0187	10.5303 ±0.0141	**5.9968** ±0.0017
SSR-FM (SDE)	2.1000 ±0.0081	**1.5091** ±0.0066	**3.0834** ±0.0049	6.2795 ±0.0104	**10.3301** ±0.0043	6.0400 ±0.0015

Table B10: EMT and CITE-seq: full-cell W_1 at each individually held-out snapshot. Means over five training seeds (0–4), with sample SD. Lower is better. Bold and underlining mark the best and second-best means before rounding.

Method	EMT (10D)		CITE-seq (50D)	
	t_1	t_2	t_1	t_2
OT-CFM	1.818 ±0.001	2.121 ±0.016	6.085 ±0.003	6.863 ±0.018
SF²M (ODE)	**1.816 ±0.002**	2.114 ±0.008	6.072 ±0.003	6.410 ±0.014
SF²M (SDE)	1.835 ±0.002	2.112 ±0.008	6.113 ±0.001	6.244 ±0.011
Metric FM	2.310 ±1.070	2.107 ±0.006	6.053 ±0.004	6.415 ±0.033
RFM	1.994 ±0.001	2.168 ±0.005	6.036 ±0.003	**5.935 ±0.019**
RDSB	2.005 ±0.020	2.270 ±0.014	6.364 ±0.005	6.019 ±0.007
SSR-FM (ODE)	1.831 ±0.003	**2.098 ±0.008**	**5.997 ±0.002**	6.184 ±0.013
SSR-FM (SDE)	1.845 ±0.002	2.100 ±0.008	6.040 ±0.001	6.064 ±0.009

Table B11: EB in 10D and 20D: full-cell W_1 at each individually held-out snapshot. Means over five training seeds (0–4), with sample SD. Lower is better. Bold and underlining mark the best and second-best means before rounding.

Method	EB (10D)			EB (20D)		
	t_1	t_2	t_3	t_1	t_2	t_3
OT-CFM	1.701 ±0.003	1.515 ±0.026	1.611 ±0.014	3.183 ±0.004	3.115 ±0.019	3.477 ±0.012
SF²M (ODE)	1.699 ±0.003	1.518 ±0.007	1.616 ±0.011	3.182 ±0.002	3.107 ±0.006	3.432 ±0.013
SF²M (SDE)	1.702 ±0.003	1.518 ±0.004	**1.610 ±0.008**	3.185 ±0.003	3.096 ±0.008	3.384 ±0.011
Metric FM	**1.679 ±0.004**	1.558 ±0.007	1.624 ±0.022	3.174 ±0.001	3.655 ±1.161	3.453 ±0.051
RFM	1.820 ±0.012	1.699 ±0.010	1.676 ±0.015	3.224 ±0.003	3.189 ±0.016	**3.322 ±0.007**
RDSB	2.275 ±0.047	3.000 ±0.010	2.668 ±0.056	3.323 ±0.010	3.645 ±0.022	3.546 ±0.031
SSR-FM (ODE)	1.682 ±0.003	1.517 ±0.002	1.639 ±0.013	**3.148 ±0.003**	3.099 ±0.006	3.368 ±0.024
SSR-FM (SDE)	1.682 ±0.003	**1.509 ±0.007**	1.632 ±0.009	3.153 ±0.002	**3.083 ±0.005**	3.340 ±0.020

at t_3. Thus the learned geometry improves several missing-time reconstructions without improving every holdout.

Table B12: EB in 50D and 100D: full-cell W_1 at each individually held-out snapshot. Means over five training seeds (0–4), with sample SD. Lower is better. Bold and underlining mark the best and second-best means before rounding.

Method	EB (50D)			EB (100D)		
	t_1	t_2	t_3	t_1	t_2	t_3
OT-CFM	5.957 ±0.005	7.140 ±0.026	9.667 ±0.153	9.139 ±0.007	11.538 ±0.041	15.808 ±0.333
SF²M (ODE)	5.939 ±0.004	6.648 ±0.017	7.812 ±0.032	9.213 ±0.004	10.670 ±0.008	12.236 ±0.017
SF²M (SDE)	6.015 ±0.003	6.411 ±0.008	7.177 ±0.017	9.453 ±0.003	10.438 ±0.004	11.569 ±0.013
Metric FM	5.955 ±0.003	6.771 ±0.023	8.242 ±0.051	9.146 ±0.005	11.239 ±0.022	13.753 ±0.086
RFM	**5.838 ±0.004**	**6.132 ±0.008**	6.773 ±0.057	**8.960 ±0.016**	10.450 ±0.069	12.724 ±0.324
RDSB	6.162 ±0.014	6.546 ±0.021	**6.753 ±0.031**	10.117 ±0.010	10.996 ±0.011	11.412 ±0.010
SSR-FM (ODE)	5.843 ±0.003	6.508 ±0.019	7.444 ±0.029	9.099 ±0.008	10.530 ±0.014	11.623 ±0.074
SSR-FM (SDE)	5.942 ±0.003	6.279 ±0.010	6.994 ±0.017	9.400 ±0.002	**10.330 ±0.004**	**11.226 ±0.043**

EB in 50D and 100D. Table B12 shows that SSR-FM (SDE) is best at EB100D t_2 and t_3, reaching $W_1 = 11.226$ at t_3 compared with 11.412 for the strongest baseline, RDSB. SSR-FM (ODE) is second at EB100D t_1. RFM is best at EB50D t_1 and t_2, with SSR-FM second through its ODE and SDE, respectively. At EB50D t_3, RDSB has the lowest mean (6.753), followed by RFM (6.773) and SSR-FM (SDE, 6.994).

Across all 16 settings, SSR-FM has lower mean W_1 than SF²M in 14 under SDE sampling and in the same 14 under ODE sampling from the same trained models. The exceptions are EMT t_1 and EB10D t_3. These results support competitive interpolation with data-estimated geometry across 10–100 dimensions, with gains depending on the representation and held-out time; they do not establish recovery of individual biological trajectories.

C Scalable Coupling and Conditional Computation

Computing a geometric path and its full action Hessian for every endpoint pair is costly for large, high-dimensional populations. The single-cell experiments therefore combine blockwise sparse coupling with a local quadratic approximation to the conditional covariance and endpoint prefactor. We describe the sparse computation, derive the local quadratic covariance and prefactor, and evaluate the approximation in a controlled EB50D comparison. The LAND metric is still spatially varying; its definition is unchanged from (B.9).

C.1 Scalable Preprocessing and Sampling

For each pair of neighbouring training snapshots, we independently shuffle the two populations and form paired minibatches with at most M samples from each population.

Within each minibatch, we select candidate source–target edges using squared coordinate distances. We keep the r highest-weight connections for each source and target in an entropic OT plan, together with the nonzero edges of an exact OT plan to preserve feasibility. Only these selected edges, denoted by $\mathcal{E}_b$, require geometric path and prefactor computations. Sparse Sinkhorn then assigns their transport probabilities using the Riemannian costs. Algorithm 2 summarizes the procedure and conditional sampling.

For a minibatch with n_b source and m_b target samples, set $a_i = 1/n_b$ and $b_j = 1/m_b$. For each selected edge, let S_{ij} be the numerically minimized Brownian action and $\widetilde{\Delta}_{ij}$ the prefactor from Section C.2. The sparse coupling solves

Algorithm 2 Scalable preprocessing and conditional sampling

Require: Source $\{x_i\}_{i=1}^{N_0}$, target $\{y_j\}_{j=1}^{N_1}$, metric G, and interval noise ε.
Require: Minibatch cap M, candidate count r, and marginal tolerance δ_{OT}.
Preprocessing:
Independently shuffle the populations and form paired minibatches of size at most M.
for each paired minibatch b **do**
  Select the r highest-weight edges per source and target from coordinate-space entropic OT.
  Add the nonzero edges of exact coordinate-space OT to obtain $\mathcal{E}_b$.
  for each $(i, j) \in \mathcal{E}_b$ **do**
    Minimize $S_{g,h}(x_i, \mathbf{x}_h, y_j)$; store the path γ_h^{ij} and action S_{ij}.
    Compute segment metrics $\overline{G}_k^{ij}$ and store cumulative A_t^{ij} using (C.2)–(C.3).
    $\widetilde{C}_{ij} \leftarrow S_{ij} - \frac{\varepsilon}{2} \log \widetilde{\Delta}_{ij}$, with $\widetilde{\Delta}_{ij}$ from (C.4).
  end for
  Solve (C.1) to tolerance δ_{OT}; store $\widehat{\pi}_b$ and $w_b = n_b/N_0$.
end for
Conditional sampling:
Sample b with probability w_b, then $z = (x_i, y_j) \sim \widehat{\pi}_b$.
Sample a local training time t and $\xi \sim \mathcal{N}(0, I)$.
Evaluate $\gamma_t(z)$, $\widetilde{\Sigma}_t(z)$, and $\partial_t \widetilde{\Sigma}_t(z)$ as in Section C.2.
$x \leftarrow \gamma_t(z) + \sqrt{\varepsilon}\, \widetilde{\Sigma}_t(z)^{1/2} \xi$.
return t, x and the flow/score targets (4.6)–(4.7), using $\widetilde{\Sigma}_t$.

$$\begin{aligned} \widetilde{C}_{ij} &= S_{ij} - \frac{\varepsilon}{2} \log \widetilde{\Delta}_{ij}, \\ \widehat{\pi}_b &= \operatorname*{arg\,min}_{\substack{\pi \mathbf{1} = a,\ \pi^\top \mathbf{1} = b,\ \pi \geq 0 \\ \operatorname{supp}(\pi) \subseteq \mathcal{E}_b}} \sum_{(i,j) \in \mathcal{E}_b} \left[\pi_{ij} \widetilde{C}_{ij} + \varepsilon \pi_{ij} (\log \pi_{ij} - 1) \right]. \end{aligned} \tag{C.1}$$

We solve (C.1) by log-domain sparse Sinkhorn scaling to a prescribed maximum absolute marginal residual δ_{OT}, using dual L-BFGS refinement when needed. Endpoint pairs are sampled according to the resulting transport probabilities, not uniformly over the selected edges.

A block containing n_b of the N_0 source cells is sampled with probability $w_b = n_b/N_0$. Thus $\widehat{\pi} = \sum_b w_b \widehat{\pi}_b$ preserves uniform source weights, while each target cell in block b has weight w_b/m_b. Integer block sizes can make the latter differ slightly from globally uniform target weights. The result is a mixture of balanced block couplings, not the full-support global entropic OT solution.

C.2 Local Quadratic Covariance and Prefactor

This section derives the covariance and endpoint prefactor used in Algorithm 2. The approximation preserves the optimized mean path while avoiding formation and inversion of the full action Hessian.

For a retained endpoint pair, we minimize the midpoint discretization of $S_g = E_g/2$ using the original action and its exact first derivatives. On the grid $0 = t_0 < \cdots < t_{K+1} = 1$, write $h_k = t_{k+1} - t_k$ and let $x^\star_{t_k,h}$ denote the resulting path. Evaluate

$$\begin{aligned} \overline{G}_k &= G\left(\frac{x^\star_{t_k,h} + x^\star_{t_{k+1},h}}{2} \right), \\ Q_h(\eta) &= \frac{1}{2} \sum_{k=0}^{K} \frac{(\eta_{k+1} - \eta_k)^\top \overline{G}_k (\eta_{k+1} - \eta_k)}{h_k}, \end{aligned} \tag{C.2}$$

where $\eta_0 = \eta_{K+1} = 0$. The matrices $\overline{G}_k$ vary along the path. Unlike the full action Hessian, Q_h omits derivatives of G with respect to the fluctuating coordinates; it does not change the action used to optimize the mean path.

Let $\overline{G}(s) = \overline{G}_k$ on $[t_k, t_{k+1})$ and define the integrated cometric

$$A_t = \int_0^t \overline{G}(s)^{-1}\,\mathrm{d}s, \qquad \widetilde{\Sigma}_{t,h} = A_t - A_t A_1^{-1} A_t. \tag{C.3}$$

To obtain (C.3), take independent Gaussian increments with covariances $\varepsilon h_k \overline{G}_k^{-1}$ and condition their sum to be zero. Gaussian conditioning gives covariance $\varepsilon\widetilde{\Sigma}_{t,h}$; at grid nodes this is exactly the corresponding block of $\varepsilon(\nabla^2 Q_h)^{-1}$. Between nodes, the same formula uses the piecewise-constant cometric and a linearly interpolated mean path. It requires only $d \times d$ matrix operations rather than inversion of the full $Kd \times Kd$ action Hessian.

The boundary Schur precision of the same quadratic form is A_1^{-1}. We therefore use the consistent endpoint prefactor

$$\widetilde{\Delta}(x_0, x_1) = \frac{\det(A_1)^{-1}}{\sqrt{|G(x_0)|\,|G(x_1)|}}. \tag{C.4}$$

For diagonal LAND, write $A_t = \mathrm{diag}(a_1(t), \ldots, a_d(t))$ and $\overline{G}(t) = \mathrm{diag}(\overline{g}_1(t), \ldots, \overline{g}_d(t))$. Then the covariance and its time derivative reduce to

$$\begin{aligned} \widetilde{\Sigma}^{jj}_{t,h} &= a_j(t)\left(1 - \frac{a_j(t)}{a_j(1)}\right), \\ \partial_t \widetilde{\Sigma}^{jj}_{t,h} &= \overline{g}_j(t)^{-1}\left(1 - \frac{2a_j(t)}{a_j(1)}\right). \end{aligned} \tag{C.5}$$

We substitute these quantities into the Gaussian sampler and flow/score targets (4.5)–(4.7). The derivative is taken within each segment; grid knots have zero probability under continuous time sampling. The Riemannian-volume score correction and the coordinate Itô drift still use the original $G(x)$. For a constant metric, (C.3) gives $\widetilde{\Sigma}_{t,h} = t(1-t)G^{-1}$ and (C.4) gives $\widetilde{\Delta} = 1$. For a varying metric, this computational approximation is distinct from the full action-Hessian construction and does not automatically inherit its small-noise guarantee. The single-cell implementation uses four equal path segments.

C.3 Controlled Approximation Study on EB50D

Setting. We evaluate the scalable construction on the EB data in 50 dimensions, withholding the middle snapshot at $t = 0.5$. The 13,541 training cells define the same LAND metric and 49 paired minibatches, with minibatch cap $M = 256$ and candidate count $r = 32$. This controlled study fixes $\beta = 0.02$ and four action segments across all variants. We vary two choices: the covariance and endpoint prefactor (Full Hessian or Quadratic), and the endpoint-pair support (full or sparse). Full Hessian uses the complete action-Hessian construction; Quadratic uses (C.3)–(C.4). Full support contains all endpoint pairs within each minibatch, not a global coupling over all cells; sparse support retains the selected pairs in Algorithm 2. All variants minimize the same action under the same metric. Both sparse variants enter the coupling comparison, while the prediction comparison includes Full Hessian with full support and Quadratic with full or sparse support.

Coupling comparison. To isolate pure sparsification, we compare the full and sparse supports separately under each fixed endpoint kernel. Let π_b^{full} and π_b^{sp} be the normalized plans in minibatch b, with zero mass assigned to omitted pairs. We use the training sampling weights: minibatch source mass w_b multiplied by its observation-interval duration Δt_b. Total variation (TV) and the probability overlap are

$$\begin{aligned} \mathrm{TV}_b &= \frac{1}{2}\sum_{i,j} |(\pi_b^{\mathrm{sp}})_{ij} - (\pi_b^{\mathrm{full}})_{ij}|, \qquad \omega_b = \frac{\Delta t_b w_b}{\sum_c \Delta t_c w_c}, \\ \mathrm{Overlap} &= \sum_b \omega_b \sum_{i,j} \min\{(\pi_b^{\mathrm{sp}})_{ij}, (\pi_b^{\mathrm{full}})_{ij}\} = 1 - \sum_b \omega_b \mathrm{TV}_b. \end{aligned} \tag{C.6}$$

The factor $1/2$ counts redistributed probability mass only once. Overlap measures the shared mass of the training endpoint-pair distributions, rather than the fraction of common edges or the coincidence rate of independent draws. It is computed directly from the plans and reported as a percentage. Each sparse plan is compared with its own full-support reference using the same kernel and minibatch partition.

Table C1: EB50D approximation study. Each row specifies the covariance/prefactor construction and pair support. Overlap compares each sparse coupling with its own full-support reference under the same kernel, as defined in (C.6). Held-out W_1 at $t = 0.5$ is the mean $\pm$ SD over five 512-cell evaluations from one matched training run per evaluated construction. The sparse Full Hessian variant is evaluated only for coupling overlap.

Covariance / prefactor	Pair support	Overlap (%) ↑	W_1 ↓
Full Hessian	Full	Reference	7.056 ± 0.053
Full Hessian	Sparse	99.85	—
Quadratic	Full	Reference	7.101 ± 0.049
Quadratic	Sparse	99.70	7.099 ± 0.047

Table C1 shows high coupling overlap (99.70%–99.85%) and similar held-out W_1 across the three constructions. These EB50D results support the approximation as a practical way to accelerate geometric preprocessing on real data through sparse edge selection and local quadratic computation, while retaining comparable interpolation accuracy.

D Ablation Studies

These controls ask three questions: when does the full conditional covariance improve interpolation, does its benefit already appear before field learning, and can the volume correction J change endpoint pairings? Tables D1 and D2 keep the coupling fixed to isolate conditional and sampling choices; Sections D.3 and D.4 instead vary the endpoint kernel. Appendix C separately examines the scalable coupling approximation.

D.1 Conditional Covariance and Deterministic Controls

On Saddle, Helicoid, IBTrACS, and EB10D, we hold the metric, mean paths, endpoint coupling, initialization, and pair/time draws fixed. We replace the full-action-Hessian covariance by either

$$\Sigma_t^{\mathrm{loc}} = t(1-t)G(\gamma_t)^{-1}, \qquad \Sigma_t^{\mathrm{quad}} = A_t - A_t A_1^{-1} A_t, \tag{D.1}$$

where A_t is the integrated cometric in (C.3), and update the flow and score targets consistently. Both covariance replacements retain SDE sampling. Geodesic FM is a different control: it is separately trained on the same coupled mean paths without conditional noise or score matching, rather than using the independent endpoint pairs of the RFM baseline. SSR-FM (ODE) integrates only the flow field of the same trained model as SSR-FM (SDE), without retraining on deterministic targets.

Each variant uses two matched 20,000-update runs. The surface tasks use $\varepsilon = 0.1$; IBTrACS and EB10D use $\beta = 0.005$. EB10D retains full action Hessians on the shared sparse support. The IBTrACS ablation uses great-circle transport plus normal deviation in km, rather than the main comparison's chord distance.

The full action-Hessian covariance gives comparable results to the simpler covariance models on Saddle, IBTrACS, and EB10D, and substantially improves interpolation on Helicoid. There, it reduces W_1 from 0.0984 (quadratic) to 0.0678, a 31.1% reduction, with the same ordering in both runs. Since the coupling and mean paths are held fixed, this improvement highlights the contribution of the conditional covariance to recovering intermediate distributions.

The full SDE also improves over separately trained Geodesic FM on both surfaces. For the same trained SSR-FM model, SDE sampling reduces Helicoid W_1 from 0.0802 (ODE) to 0.0678. Taken together with the results in Sections 6.2 and 6.3, these comparisons show an overall benefit from

Table D1: Controlled field-training ablations. Hidden-time W_1, averaged over two training runs, with sample SD. All rows use the same endpoint coupling and mean paths. Lower is better; bold and underlining indicate the best and second-best means before rounding.

Conditional / sampler	Saddle	Helicoid	IBTrACS (km)	EB10D
Local-metric covariance	0.1122 ± 0.0051	0.0993 ± 0.0045	119.644 ± 0.712	1.5171 ± 0.0091
Quadratic covariance	$\mathbf{0.1106 \pm 0.0050}$	0.0984 ± 0.0057	119.644 ± 0.712	1.5167 ± 0.0107
Geodesic FM	0.1231 ± 0.0002	$\underline{0.0798 \pm 0.0012}$	$\underline{119.200 \pm 0.865}$	$\mathbf{1.5157 \pm 0.0013}$
SSR-FM (ODE)	0.1149 ± 0.0002	0.0802 ± 0.0009	$\mathbf{119.190 \pm 0.785}$	1.5243 ± 0.0105
SSR-FM (SDE)	$\underline{0.1110 \pm 0.0042}$	$\mathbf{0.0678 \pm 0.0019}$	119.644 ± 0.712	$\underline{1.5160 \pm 0.0102}$

stochastic SDE reconstruction: it improves distribution matching on both molecular datasets and all four EB representations, although ODE sampling gives slightly lower errors on some datasets.

D.2 Conditional Accuracy before Field Learning

We assess the contribution of the conditional construction to the Helicoid gain independently of field learning. We directly draw midpoint samples from the same coupling and mean paths using the full, local-metric, or quadratic covariance; the geodesic-mean control removes the fluctuation term. No neural network is used. The two surfaces use $\varepsilon = 0.1$ and 32 action segments, and the numerical Brownian reference is matched to the same 256 empirical endpoints per distribution. We compare raw-embedding W_1. As a sampling control, we also compute W_1 between two independent sample sets drawn from the same numerical Brownian reference.

Table D2: Direct midpoint sampling without a neural network. W_1 to the numerical Brownian reference with matched empirical endpoints; mean and sample SD over two Monte Carlo seeds, each averaging five 512-sample comparisons. Bold and underlining rank only the four direct constructions. The last row compares two independent sample sets from the same reference.

Construction	Saddle	Helicoid
Local-metric covariance	0.0815 ± 0.0010	0.0947 ± 0.0013
Quadratic covariance	$\mathbf{0.0807 \pm 0.0008}$	0.0916 ± 0.0011
Geodesic mean only	0.0938 ± 0.0025	$\underline{0.0837 \pm 0.0004}$
Full action Hessian	$\underline{0.0811 \pm 0.0011}$	$\mathbf{0.0640 \pm 0.0004}$
Independent reference samples	0.0796 ± 0.0022	0.0403 ± 0.0017

Direct sampling with the full covariance reduces Helicoid W_1 from 0.0916 (quadratic) to 0.0640, a 30.1% reduction, with the same ordering in both Monte Carlo seeds. On Saddle, the full and quadratic covariances give similar values (0.0811 and 0.0807), close to the 0.0796 obtained from two independent sets of reference samples. The agreement between Tables D1 and D2 shows that the Helicoid improvement originates in the conditional construction and is retained after field learning.

D.3 Volume Correction and Endpoint Pairing

Endpoint marginal matching alone cannot reveal the effect of J, since balanced couplings share those marginals. We therefore compare the pairings induced by the distance-only kernel $K_{ij}^{\mathrm{dist}} = \exp[-d_g(x_i, y_j)^2/(2\varepsilon)]$, the corrected kernel $K_{ij}^{J} = K_{ij}^{\mathrm{dist}} J(x_i, y_j)^{-1/2}$, and an independent Brownian heat kernel. Geometry, endpoints, marginal weights, and noise are identical, and no neural field is trained. The diagnostics are conditional destination distributions and full-plan $\mathrm{TV}(\pi, \pi^{\mathrm{heat}}) = \frac{1}{2}\sum_{ij} |\pi_{ij} - \pi_{ij}^{\mathrm{heat}}|$.

As in Appendix A.4, take orthonormal tangent coordinates η at x and set $y = \mathrm{Exp}_x(\eta)$. Then

$$J(x,y) = \sqrt{|G(y)|}\,|\det D_\eta \operatorname{Exp}_x(\eta)|\,, \qquad \mathrm{d}\mu_g(y) = J(x,y)\,\mathrm{d}\eta. \tag{D.2}$$

Thus J measures volume expansion of the exponential map, not an arbitrary coordinate Jacobian; the Van Vleck–Morette determinant is J^{-1}. Distance determines the minimum action, whereas $J^{-1/2}$ accounts for the leading effect of transverse geometric fluctuations on transition density. At equal action, expansion ($J > 1$) lowers the heat-kernel prefactor and compression ($J < 1$) raises it. This is the classical heat-kernel correction used in (4.1), not an additional fitted parameter.

The correction matters only through its nonseparable endpoint dependence. Positive Sinkhorn row and column scalings cancel from every 2×2 cross ratio, giving

$$\log \frac{\widehat{\pi}^J_{11}\widehat{\pi}^J_{22}}{\widehat{\pi}^J_{12}\widehat{\pi}^J_{21}} = -\frac{e_{11}+e_{22}-e_{12}-e_{21}}{\varepsilon} - \frac{1}{2}\log\frac{J_{11}J_{22}}{J_{12}J_{21}}, \qquad e_{ij} = \frac{1}{2}d_g(x_i,y_j)^2. \tag{D.3}$$

If $\log J(x_i, y_j) = a_i + b_j$, the correction is absorbed by these scalings and does not change the balanced coupling. Otherwise it can reverse a pairing preference when the competing actions are close. It need not dominate a fixed, well-separated action gap as $\varepsilon \downarrow 0$.

The ring torus has major radius 1.15, minor radius 1, and metric $G = \operatorname{diag}(1, (1.15 + \cos\theta)^2)$ in minor/major angular coordinates (θ, ϕ). At $\varepsilon = 0.5$, we use

$$\begin{aligned} &x_1 = (0, -a), && x_2 = (2.8, 1.8 + a), \\ &y_1 = (2.8, 1.8 - a), && y_2 = (0, a), \qquad a = 1.0069977511350234. \end{aligned} \tag{D.4}$$

Figure D1 visualizes the endpoint masses and connecting routes alongside the conditional destination probabilities. The reference comes from an independent Fourier-spectral Brownian heat-kernel calculation. Including J changes $P(y_2 \mid x_1)$ from 34.40% to 66.69%, close to the reference 66.29%, and reduces full-plan TV from 0.127565 to 0.001616. Thus the correction recovers a pairing preference that distance alone reverses; it is not merely a normalization that disappears under balancing. This is a constructed example selected to expose the correction: the outer-equator path reaches 94% of its first conjugate length, so it does not establish uniform accuracy near conjugate points.

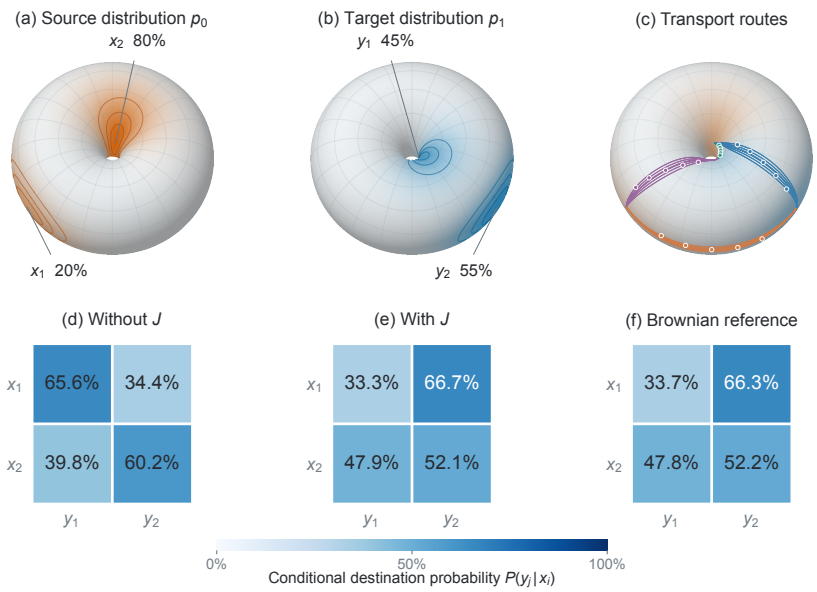


Figure D1: Endpoint distributions and transport routes for the ring-torus J control. (a,b) Source and target masses, broadened into coloured patches with contours for display only; the couplings use the original atomic endpoints. (c) Saved endpoint geodesics with narrow deterministic curve bundles and particle markers for illustration. Colours distinguish endpoint pairs; no stochastic fluctuations are shown. (d–f) Conditional destination probabilities $P(y_j \mid x_i) = \pi_{ij}/p_0(\{x_i\})$ from the original plans, on a common 0–100% colour scale.

D.4 Analytic Verification of the Volume Correction

An analytic counterpart removes the need for a numerical heat-kernel reference. We use the four-dimensional product of hyperbolic three-space with curvature -1 and a flat line. In Poincaré ball coordinates, the manifold and metric are

$$\begin{aligned} \mathcal{M} &= \mathbb{H}^3_{-1} \times \mathbb{R} = \{(u, z) \in \mathbb{R}^3 \times \mathbb{R} : \|u\| < 1\}, \\ G(u, z) &= \operatorname{diag}\left(\frac{4}{(1 - \|u\|^2)^2} I_3,\ 1 \right). \end{aligned} \tag{D.5}$$

Let $h(s) = (\tanh(s/2), 0, 0)$ be a unit-speed hyperbolic geodesic and $F(s, z) = (h(s), z)$. The source and target distributions each put equal mass on two opposite corners of a rectangle:

$$\begin{aligned} x_1 &= F(-1.5, -1.51), & x_2 &= F(1.5, 1.51), \\ y_1 &= F(-1.5, 1.51), & y_2 &= F(1.5, -1.51), \\ p_0 &= \tfrac{1}{2}(\delta_{x_1} + \delta_{x_2}), & p_1 &= \tfrac{1}{2}(\delta_{y_1} + \delta_{y_2}). \end{aligned} \tag{D.6}$$

The section $F(\mathbb{R}^2)$ has induced metric $\mathrm{d}s^2 + \mathrm{d}z^2$, but the Brownian reference evolves on the full four-dimensional manifold. Same-index pairs therefore differ by 3.02 along the flat factor, whereas cross-index pairs differ by 3 along the hyperbolic factor.

We fix $\varepsilon = 0.1$. For $x = (u, z)$, $y = (v, w)$, and $r = d_{\mathbb{H}}(u, v)$, the volume correction and exact transition density are

$$J(x, y) = \left(\frac{\sinh r}{r} \right)^2, \qquad k^g_\varepsilon(x, y) = \frac{e^{-\varepsilon/2}}{(2\pi\varepsilon)^2} \exp\left[-\frac{r^2 + (z - w)^2}{2\varepsilon} \right] \frac{r}{\sinh r}, \tag{D.7}$$

with both ratios defined by continuity at $r = 0$. The heat kernel follows by multiplying the hyperbolic heat kernel (Vázquez, 2022) by the one-dimensional Euclidean heat kernel. Hence K^J and the exact kernel differ only by an endpoint-independent factor and give the same balanced coupling.

For equal two-point marginal weights and a symmetric kernel $K = \left(\begin{smallmatrix} A & B \\ B & A \end{smallmatrix}\right)$, both row sums equal $A + B$, so $\widehat{\pi} = K/[2(A + B)]$. This gives the analytic probabilities in Table D3 without an iterative solver.

Table D3: Analytic hyperbolic-product coupling. Each row reports conditional probabilities from x_1; symmetry gives the other source row by exchanging the destinations. TV uses the full balanced plan. Values are analytic, with no training or Monte Carlo error.

Kernel	$P(y_1 \mid x_1)$	$P(y_2 \mid x_1)$	TV to exact heat
Without J	35.3886%	64.6114%	0.292630
With J	64.6516%	35.3484%	0
Exact heat kernel	64.6516%	35.3484%	0

Although the hyperbolic pairing is slightly shorter, volume expansion makes the flat-direction pairing more probable under the Brownian reference. The correction recovers this preference exactly in this analytic control, not necessarily on general manifolds.

E Training Efficiency

We compare SSR-FM and RDSB on the single-cell tasks in Section 6.3, using one NVIDIA RTX 4090D per run and batch size 512. Figure E1(a) reports five seeds of the original recipes: 20,000 joint flow/score updates for SSR-FM and approximately 50,000 total updates for RDSB, with 10 simulation steps and two IPF rounds. SSR-FM uses the sparse-coupling, frozen-LAND approximation in Appendix C. Training time includes RDSB's trajectory simulations but excludes preprocessing, loading, prediction generation, and evaluation. Together with the lower mean held-out W_1 in Table 3, these timings show a favorable training-time–accuracy trade-off under the reported recipes.

Figure E1(b) compares EB100D runs with one fresh seed per method, three 512-unit hidden layers, and separate GPUs. SSR-FM completes 200,000 updates in 16.89 minutes; RDSB takes 119.66 minutes for 199,992 updates. Avoiding repeated SDE simulation reduces field-training time 7.08-fold under these method-specific update budgets.

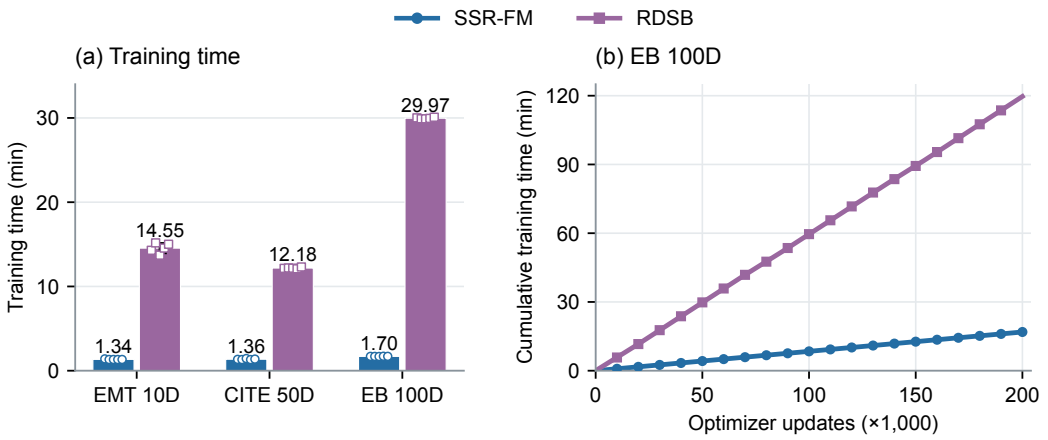


Figure E1: Training efficiency. (a) Five-seed mean and sample SD; markers show individual runs. (b) Single-seed EB100D traces with 10k-update markers interpolated between logs where necessary.

With the fitted metric and endpoint proposals, SSR-FM preprocessing adds 3.62 minutes. These field-training timings do not measure full-Hessian scaling or prediction accuracy at 200k updates.

F Comparison with Related Methods

Table F1: Method comparison. Stochasticity denotes SDE dynamics; coupling denotes an endpoint law optimized through OT or SB. Geometry-aware refers to conditional dynamics.

Method	Stochasticity	Coupling	Geometry-aware	Simulation-free training
OT-CFM	×	✓	×	✓
SF^{2}M	✓	✓	×	✓
MFM	×	✓	✓	✓
RFM	×	×	✓	✓
RDSB	✓	✓	✓	×
SSR-FM	✓	✓	✓	✓

MFM denotes OT-MFM here; I-MFM uses independent pairing. RFM is simulation-free with analytic geodesics. SF^{2}M-Geo changes the coupling but retains Euclidean conditionals. RDSB optimizes its coupling implicitly through IPF. SSR-FM precomputes geometry before field training.

Relation with other work. OT-CFM uses Euclidean OT and straight interpolants (Tong et al., 2024a), while SF²M adds Brownian conditionals and score matching (Tong et al., 2024b). MFM and RFM incorporate geometry into deterministic interpolants (Kapuśniak et al., 2024; Chen & Lipman, 2024). RDSB combines manifold geometry and stochastic dynamics through iterative diffusion simulation (Thornton et al., 2022). SSR-FM brings these properties together: a common Riemannian Brownian reference yields a geometric coupling and a Gaussian conditional probability path with geodesic mean and action-Hessian covariance. These supply explicit flow and score targets for simulation-free training in the regular small-noise setting of Section 4.